

**NOVEL POLYPEPTIDES AND NUCLEIC ACIDS  
ENCODING SAME**

**RELATED APPLICATIONS**

This application claims priority to USSN 60/177,839, filed January 25, 2000; USSN 60/176,134, filed January 14, 2000; USSN 60/175,989, filed January 13, 2000; USSN 60/218,324, filed July 14, 2000; USSN 60/220,253, filed July 24, 2000; USSN 60/178,191, filed January 26, 2000; USSN 60/178,227, filed January 26, 2000; USSN 60/220,590, filed July 25, 2000, USSN 60/215,855 (21402-048) filed July 3, 2000, and USSN 09/761,288 (15966-638 utility) filed January 16, 2001, which are incorporated herein by reference in their entirety.

**TECHNICAL FIELD OF THE INVENTION**

The invention generally relates to nucleic acids and polypeptides encoded therefrom.

**BACKGROUND OF THE INVENTION**

Within the animal kingdom, odor detection is a universal tool used for social interaction, predation, and reproduction. Chemosensitivity in vertebrates is modulated by bipolar sensory neurons located in the olfactory epithelium, which extend a single, highly arborized dendrite into the mucosa while projecting axons to relay neurons within the olfactory bulb. The many ciliae on the neurons bear odorant (or olfactory) receptors (ORs), which cause depolarization and formation of action potentials upon contact with specific odorants. ORs may also function as axonal guidance molecules, a necessary function as the sensory neurons are normally renewed continuously through adulthood by underlying populations of basal cells.

The mammalian olfactory system is able to distinguish several thousand odorant molecules. Odorant receptors are believed to be encoded by an extremely large subfamily of G protein-coupled receptors. These receptors share a 7-transmembrane domain structure with many neurotransmitter and hormone receptors and are likely to underlie the recognition and G-protein-mediated transduction of odorant signals and possibly other chemosensing responses as well. The genes encoding these receptors are devoid of introns within their coding regions. Schurmans and co-workers cloned a member of this family of genes, OLFR1, from a genomic library by cross-hybridization with a gene fragment obtained by PCR. See *Schurmans et al.*, Cytogenet.

Cell Genet., 1993, 63(3):200. By isotopic *in situ* hybridization, they mapped the gene to 17p13-p12 with a peak at band 17p13. A minor peak was detected on chromosome 3, with a maximum in the region 3q13-q21. After MspI digestion, a restriction fragment length polymorphism (RFLP) was demonstrated. Using this in a study of 3 CEPH pedigrees, they demonstrated linkage with D17S126 at 17pter-p12; maximum lod = 3.6 at theta = 0.0. Used as a probe on Southern blots under moderately stringent conditions, the cDNA hybridized to at least 3 closely related genes. Ben-Arie and colleagues cloned 16 human OLFR genes, all from 17p13.3. See Ben-Arie *et al.*, Hum. Mol. Genet., 1994, 3(2):229. The intronless coding regions are mapped to a 350-kb contiguous cluster, with an average intergenic separation of 15 kb. The OLFR genes in the cluster belong to 4 different gene subfamilies, displaying as much sequence variability as any randomly selected group of OLFRs. This suggested that the cluster may be one of several copies of an ancestral OLFR gene repertoire whose existence may have predated the divergence of mammals. Localization to 17p13.3 was performed by fluorescence *in situ* hybridization as well as by somatic cell hybrid mapping.

Previously, OR genes cloned in different species were from disparate locations in the respective genomes. The human OR genes, on the other hand, lack introns and may be segregated into four different gene subfamilies, displaying great sequence variability. These genes are primarily expressed in olfactory epithelium, but may be found in other chemoresponsive cells and tissues as well.

Blache and co-workers used polymerase chain reaction (PCR) to clone an intronless cDNA encoding a new member (named OL2) of the G protein-coupled receptor superfamily. See Blache *et al.*, Biochem. Biophys. Res. Commun., 1998, 242(3):669. The coding region of the rat OL2 receptor gene predicts a seven transmembrane domain receptor of 315 amino acids. OL2 has 46.4 percent amino acid identity with OL1, an olfactory receptor expressed in the developing rat heart, and slightly lower percent identities with several other olfactory receptors. PCR analysis reveals that the transcript is present mainly in the rat spleen and in a mouse insulin-secreting cell line (MIN6). No correlation was found between the tissue distribution of OL2 and that of the olfaction-related GTP-binding protein Golf alpha subunit. These findings suggest a role for this new hypothetical G-protein coupled receptor and for its still unknown ligand in the spleen and in the insulin-secreting beta cells.

Olfactory loss may be induced by trauma or by neoplastic growths in the olfactory neuroepithelium. There is currently no treatment available that effectively restores olfaction in

the case of sensorineural olfactory losses. See Harrison's Principles of Internal Medicine, 14<sup>th</sup> Ed., Fauci, AS *et al.* (eds.), McGraw-Hill, New York, 1998, 173. There thus remains a need for effective treatment to restore olfaction in pathologies related to neural olfactory loss.

## SUMMARY OF THE INVENTION

The invention is based, in part, upon the discovery of novel polynucleotide sequences encoding novel polypeptides.

Accordingly, in one aspect, the invention provides an isolated nucleic acid molecule that includes the sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23 or 25 or a fragment, homolog, analog or derivative thereof. The nucleic acid can include, *e.g.*, a nucleic acid sequence encoding a polypeptide at least 85% identical to a polypeptide that includes the amino acid sequences of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24 or 26. The nucleic acid can be, *e.g.*, a genomic DNA fragment, or a cDNA molecule.

Also included in the invention is a vector containing one or more of the nucleic acids described herein, and a cell containing the vectors or nucleic acids described herein.

The invention is also directed to host cells transformed with a vector comprising any of the nucleic acid molecules described above.

In another aspect, the invention includes a pharmaceutical composition that includes a NOVX nucleic acid and a pharmaceutically acceptable carrier or diluent.

In a further aspect, the invention includes a substantially purified NOVX polypeptide, *e.g.*, any of the NOVX polypeptides encoded by an NOVX nucleic acid, and fragments, homologs, analogs, and derivatives thereof. The invention also includes a pharmaceutical composition that includes an NOVX polypeptide and a pharmaceutically acceptable carrier or diluent.

In still a further aspect, the invention provides an antibody that binds specifically to an NOVX polypeptide. The antibody can be, *e.g.*, a monoclonal or polyclonal antibody, and fragments, homologs, analogs, and derivatives thereof. The invention also includes a pharmaceutical composition including NOVX antibody and a pharmaceutically acceptable carrier or diluent. The invention is also directed to isolated antibodies that bind to an epitope on a polypeptide encoded by any of the nucleic acid molecules described above.

The invention also includes kits comprising any of the pharmaceutical compositions described above.

The invention further provides a method for producing an NOVX polypeptide by providing a cell containing an NOVX nucleic acid, *e.g.*, a vector that includes an NOVX nucleic acid, and culturing the cell under conditions sufficient to express the NOVX polypeptide encoded by the nucleic acid. The expressed NOVX polypeptide is then recovered from the cell.

- 5 Preferably, the cell produces little or no endogenous NOVX polypeptide. The cell can be, *e.g.*, a prokaryotic cell or eukaryotic cell.

The invention is also directed to methods of identifying an NOVX polypeptide or nucleic acid in a sample by contacting the sample with a compound that specifically binds to the polypeptide or nucleic acid, and detecting complex formation, if present.

- 10 The invention further provides methods of identifying a compound that modulates the activity of an NOVX polypeptide by contacting an NOVX polypeptide with a compound and determining whether the NOVX polypeptide activity is modified.

- The invention is also directed to compounds that modulate NOVX polypeptide activity identified by contacting an NOVX polypeptide with the compound and determining whether the  
15 compound modifies activity of the NOVX polypeptide, binds to the NOVX polypeptide, or binds to a nucleic acid molecule encoding an NOVX polypeptide.

- In another aspect, the invention provides a method of determining the presence of or predisposition of an NOVX-associated disorder in a subject. The method includes providing a sample from the subject and measuring the amount of NOVX polypeptide in the subject sample.  
20 The amount of NOVX polypeptide in the subject sample is then compared to the amount of NOVX polypeptide in a control sample. An alteration in the amount of NOVX polypeptide in the subject protein sample relative to the amount of NOVX polypeptide in the control protein sample indicates the subject has a tissue proliferation-associated condition. A control sample is preferably taken from a matched individual, *i.e.*, an individual of similar age, sex, or other  
25 general condition but who is not suspected of having a tissue proliferation-associated condition. Alternatively, the control sample may be taken from the subject at a time when the subject is not suspected of having a tissue proliferation-associated disorder. In some embodiments, the NOVX is detected using an NOVX antibody.

- In a further aspect, the invention provides a method of determining the presence of or  
30 predisposition of an NOVX-associated disorder in a subject. The method includes providing a nucleic acid sample, *e.g.*, RNA or DNA, or both, from the subject and measuring the amount of the NOVX nucleic acid in the subject nucleic acid sample. The amount of NOVX nucleic acid



sample in the subject nucleic acid is then compared to the amount of an NOVX nucleic acid in a control sample. An alteration in the amount of NOVX nucleic acid in the sample relative to the amount of NOVX in the control sample indicates the subject has a NOVX-associated disorder.

In a still further aspect, the invention provides a method of treating or preventing or  
5 delaying an NOVX-associated disorder. The method includes administering to a subject in which such treatment or prevention or delay is desired an NOVX nucleic acid, an NOVX polypeptide, or an NOVX antibody in an amount sufficient to treat, prevent, or delay a NOVX-associated disorder in the subject.

Unless otherwise defined, all technical and scientific terms used herein have the same  
10 meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned  
15 herein are incorporated by reference in their entirety. In the case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

## 20 DETAILED DESCRIPTION OF THE INVENTION

Olfactory receptors (ORs) are the largest family of G-protein-coupled receptors (GPCRs) and belong to the first family (Class A) of GPCRs, along with catecholamine receptors and opsins. The OR family contains over 1,000 members that traverse the phylogenetic spectrum from *C. elegans* to mammals. ORs most likely emerged from prototypic GPCRs several times  
25 independently, extending the structural diversity necessary both within and between species in order to differentiate the multitude of ligands. Individual olfactory sensory neurons are predicted to express a single, or at most a few, ORs. All ORs are believed to contain seven  $\alpha$ -helices separated by three extracellular and three cytoplasmic loops, with an extracellular amino-terminus and a cytoplasmic carboxy-terminus. The pocket of OR ligand binding is expected to be  
30 between the second and sixth transmembrane domains of the proteins. Overall amino acid sequence identity within the mammalian OR family ranges from 45% to >80%, and genes greater

than 80% identical to one another at the amino acid level are considered to belong to the same subfamily.

Since the first ORs were cloned in 1991, outstanding progress has been made into their mechanisms of action and potential dysregulation during disease and disorder. It is understood that some human diseases result from rare mutations within GPCRs. Drug discovery avenues could be used to produce highly specific compounds on the basis of minute structural differences of OR subtypes, which are now being appreciated with *in vivo* manipulation of OR levels in transgenic and knock-out animals. Furthermore, due to the intracellular homogeneity and ligand specificity of ORs, renewal of specific odorant-sensing neurons lost in disease or disorder is possible by the introduction of individual ORs into basal cells. Additionally, new therapeutic strategies may be elucidated by further study of so-called orphan receptors, whose ligand(s) remain to be discovered.

OR proteins bind odorant ligands and transmit a G-protein-mediated intracellular signal, resulting in generation of an action potential. The accumulation of DNA sequences of hundreds of OR genes provides an opportunity to predict features related to their structure, function and evolutionary diversification. See Pilpel Y, et.al., Essays Biochem 1998;33:93-104. The OR repertoire has evolved a variable ligand-binding site that ascertains recognition of multiple odorants, coupled to constant regions that mediate the cAMP-mediated signal transduction. The cellular second messenger underlies the responses to diverse odorants through the direct gating of olfactory-specific cation channels. This situation necessitates a mechanism of cellular exclusion, whereby each sensory neuron expresses only one receptor type, which in turn influences axonal projections. A 'synaptic image' of the OR repertoire thus encodes the detected odorant in the central nervous system.

The ability to distinguish different odors depends on a large number of different odorant receptors (ORs). ORs are expressed by nasal olfactory sensory neurons, and each neuron expresses only 1 allele of a single OR gene. In the nose, different sets of ORs are expressed in distinct spatial zones. Neurons that express the same OR gene are located in the same zone; however, in that zone they are randomly interspersed with neurons expressing other ORs. When the cell chooses an OR gene for expression, it may be restricted to a specific zonal gene set, but it may select from that set by a stochastic mechanism. Proposed models of OR gene choice fall into 2 classes: locus-dependent and locus-independent. Locus-dependent models posit that OR genes

are clustered in the genome, perhaps with members of different zonal gene sets clustered at distinct loci. In contrast, locus-independent models do not require that OR genes be clustered.

OR genes have been mapped to 11 different regions on 7 chromosomes. These loci lie within paralogous chromosomal regions that appear to have arisen by duplications of large chromosomal domains followed by extensive gene duplication and divergence. Studies have shown that OR genes expressed in the same zone map to numerous loci; moreover, a single locus can contain genes expressed in different zones. These findings raised the possibility that OR gene choice is locus-independent or involved consecutive stochastic choices.

Issel-Tarver and Rine (1996) characterized 4 members of the canine olfactory receptor gene family. The 4 subfamilies comprised genes expressed exclusively in olfactory epithelium. Analysis of large DNA fragments using Southern blots of pulsed field gels indicated that subfamily members were clustered together, and that two of the subfamilies were closely linked in the dog genome. Analysis of the four olfactory receptor gene subfamilies in 26 breeds of dog provided evidence that the number of genes per subfamily was stable in spite of differential selection on the basis of olfactory acuity in scent hounds, sight hounds, and toy breeds.

Issel-Tarver and Rine (1997) performed a comparative study of four subfamilies of olfactory receptor genes first identified in the dog to assess changes in the gene family during mammalian evolution, and to begin linking the dog genetic map to that of humans. These four families were designated by them OLF1, OLF2, OLF3, and OLF4 in the canine genome. The subfamilies represented by these four genes range in size from 2 to 20 genes. They are all expressed in canine olfactory epithelium but were not detectably expressed in canine lung, liver, ovary, spleen, testis, or tongue. The OLF1 and OLF2 subfamilies are tightly linked in the dog genome and also in the human genome. The smallest family is represented by the canine OLF1 gene. Using dog gene probes individually to hybridize to Southern blots of genomic DNA from 24 somatic cell hybrid lines. They showed that the human homologous OLF1 subfamily maps to human chromosome 11. The human gene with the strongest similarity to the canine OLF2 gene also mapped to chromosome 11. Both members of the human subfamily that hybridized to canine OLF3 were located on chromosome 7. It was difficult to determine to which chromosome or chromosomes the human genes that hybridized to the canine OLF4 probe mapped. This subfamily is large in mouse and hamster as well as human, so the rodent background largely obscured the human cross-hybridizing bands. It was possible, however, to discern some human-specific bands in blots corresponding to human chromosome 19. They refined the mapping of the

human OLF1 homolog by hybridization to YACs that map to 11q11. In dogs, the OLF1 and OLF2 subfamilies are within 45 kb of one another (Issel-Tarver and Rine (1996)).

Issel-Tarver and Rine (1997) demonstrated that in the human OLF1 and OLF2 homologs are likewise closely linked. By studying YACs, Issel-Tarver and Rine (1997) found that the human OLF3 homolog maps to 7q35. A chromosome 19-specific cosmid library was screened by hybridization with the canine OLF4 gene probe, and clones that hybridized strongly to the probe even at high stringency were localized to 19p13.1 and 19p13.2. These clones accounted, however, for a small fraction of the homologous human bands.

Rouquier et al. (1998) demonstrated that members of the olfactory receptor gene family are distributed on all but a few human chromosomes. Through fluorescence *in situ* hybridization analysis, they showed that OR sequences reside at more than 25 locations in the human genome. Their distribution was biased for terminal bands of chromosome arms. Flow-sorted chromosomes were used to isolate 87 OR sequences derived from 16 chromosomes. Their sequence relationships indicated the inter- and intrachromosomal duplications responsible for OR family expansion. Rouquier et al. (1998) determined that the human genome has accumulated a striking number of dysfunctional copies: 72% of these sequences were found to be pseudogenes. ORF-containing sequences predominate on chromosomes 7, 16, and 17.

Trask et al. (1998) characterized a subtelomeric DNA duplication that provided insight into the variability, complexity, and evolutionary history of that unusual region of the human genome, the telomere. Using a DNA segment cloned from chromosome 19, they demonstrated that the blocks of DNA sequence shared by different chromosomes can be very large and highly similar. Three chromosomes appeared to have contained the sequence before humans migrated around the world. In contrast to its multicopy distribution in humans, this subtelomeric block maps predominantly to a single locus in chimpanzee and gorilla, that site being nonorthologous to any of the locations in the human genome. Three new members of the olfactory receptor (OR) gene family were found to be duplicated within this large segment of DNA, which was found to be present at 3q, 15q, and 19p in each of 45 unrelated humans sampled from various populations. From its sequence, one of the OR genes in this duplicated block appeared to be potentially functional. The findings raised the possibility that functional diversity in the OR family is generated in part through duplications and interchromosomal rearrangements of the DNA near human telomeres.

Mombaerts (1999) reviewed the molecular biology of the odorant receptor (OR) genes in vertebrates. Buck and Axel (1991) discovered this large family of genes encoding putative odorant receptor genes. Zhao et al. (1998) provided functional proof that one OR gene encodes a receptor for odorants. The isolation of OR genes from the rat by Buck and Axel (1991) was based on three assumptions. First, ORs are likely G protein-coupled receptors, which characteristically are 7-transmembrane proteins. Second, ORs are likely members of a multigene family of considerable size, because an immense number of chemicals with vastly different structures can be detected and discriminated by the vertebrate olfactory system. Third, ORs are likely expressed selectively in olfactory sensory neurons. Ben-Arie et al. (1994) focused attention on a cluster of human OR genes on 17p, to which the first human OR gene, OR1D2, had been mapped by Schurmans et al. (1993). According to Mombaerts (1999), the sequences of more than 150 human OR clones had been reported.

The human OR genes differ markedly from their counterparts in other species by their high frequency of pseudogenes, except the testicular OR genes. Research showed that individual olfactory sensory neurons express a small subset of the OR repertoire. In rat and mouse, axons of neurons expressing the same OR converge onto defined glomeruli in the olfactory bulb.

The present invention provides novel nucleotides and polypeptides encoded thereby. Included in the invention are the novel nucleic acid sequences and their polypeptides. The sequences are collectively referred to as "NOVX nucleic acids" or "NOVX polynucleotides" and the corresponding encoded polypeptides are referred to as "NOVX polypeptides" or "NOVX proteins." Unless indicated otherwise, "NOVX" is meant to refer to any of the novel sequences disclosed herein. Table 1 provides a summary of the NOVX nucleic acids and their encoded polypeptides. Example 1 provides a description of how the novel nucleic acids were identified.

**TABLE 1. Sequences and Corresponding SEQ ID Numbers**

NOVX Assignment	Internal Identification	SEQ ID NO (nucleic acid)	SEQ ID NO (polypeptide)	Homology
1	AL121944 A	1	2	OR GPCR
2	AL135904 A	3	4	OR GPCR
3	AL121986A	5	6	OR GPCR
4	AL121986A1	7	8	OR GPCR
5	AC012661 A	9	10	OR GPCR
6	AC012661 B	11	12	OR GPCR
7	AF061779 A	13	14	OR GPCR
8	AC012616 A	15	16	OR GPCR
9	AC012616 A1	17	18	OR GPCR
10	AC019108 A	19	20	OR GPCR
11	AC012661_dal	21	22	OR GPCR
12	CG50381-01	23	24	OR GPCR
13	AC012661A_0.4 6_EXT	25	26	OR GPCR

Where OR GPCR is an odorant receptor of the G-protein coupled-receptor family.

NOVX nucleic acids and their encoded polypeptides are useful in a variety of applications and contexts. The various NOVX nucleic acids and polypeptides according to the invention are useful as novel members of the protein families according to the presence of domains and sequence relatedness to previously described proteins. Additionally, NOVX nucleic acids and polypeptides can also be used to identify proteins that are members of the family to which the NOVX polypeptides belong.

For example, NOV1-10 are homologous to members of the odorant receptor (OR) family of the human G-protein coupled receptor (GPCR) superfamily of proteins, as shown in Table 56. Thus, the NOV1-10 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications in disorders of olfactory loss, *e.g.*, trauma, HIV illness, neoplastic growth and neurological disorders *e.g.* Parkinson's disease and Alzheimer's disease.

The NOVX nucleic acids and polypeptides can also be used to screen for molecules, which inhibit or enhance NOVX activity or function. Specifically, the nucleic acids and polypeptides according to the invention may be used as targets for the identification of small molecules that modulate or inhibit, *e.g.*, neurogenesis, cell differentiation, cell motility, cell proliferation and angiogenesis.

Additional utilities for the NOVX nucleic acids and polypeptides according to the invention are disclosed herein.



## NOVI

A NOVI sequence according to the invention is a nucleic acid sequence encoding a polypeptide related to the human odorant receptor (OR) family of the G-protein coupled receptor (GPCR) superfamily of proteins. A NOVI nucleic acid and its encoded polypeptide includes the sequences shown in Table 2. The disclosed nucleic acid (SEQ ID NO:1) is 1,071 nucleotides in length and contains an open reading frame (ORF) that begins with an ATG initiation codon at nucleotides 42-44 and ends with a TAA stop codon at nucleotides 1,053-1,055. The representative ORF encodes a 337 amino acid polypeptide (SEQ ID NO:2). Putative untranslated regions upstream and downstream of the coding sequence are underlined in SEQ ID NO: 1.

**TABLE 2.**

ATATTTCAATTCTCTGGGTCTTCATGCAGATATATTCAAGCAATGGAAGGGGAAAAATC  
AAACCAATATCTCTGAATTTCTCCTCCTGGGCTTCTCAAGTTGGCAACAACAGCAGG  
TGCTACTCTTTGCACTTTTCCTGTGTCTCTATTTAACAGGGCTGTTTGGAAACTTACT  
15 CATCTTGCTGGCCATTGGCTCGGATCACTGCCTTCACACACCCATGTATTTCTTCCTT  
GCCAATCTGTCCTTGGTAGACCTCTGCCTTCCCTCAGCCACAGTCCCCAAGATGCTA  
CTGAACATCCAAACCCAAACCCAAACCATCTCCTATCCCGGCTGCCTGGCTCAGATG  
TATTTCTGTATGATGTTTGCCAATATGGACAATTTTCTTCTCACAGTGATGGCATATG  
ACCGTTACGTGGCCATCTGTCACCCTTTACATTACTCCACCATTATGGCCCTGCGCCT  
20 CTGTGCCTCTCTGGTAGCTGCACCTTGGGTCATTGCCATTTTGAACCCTCTCTTGCAC  
ACTCTTATGATGGCCCATCTGCACTTCTGCTCTGATAATGTTATCCACCATTTCCTTCT  
GTGATATCAACTCTCTCCTCCCTCTGTCCTGTTCCGACACCAGTCTTAATCAGTTGAG  
TGTTCTGGCTACGGTGGGGCTGATCTTTGTGGTACCTTCAGTGTGTATCCTGGTATCC  
TATATCCTCATTGTTTCTGCTGTGATGAAAGTCCCTTCTGCCCAAGGAAAACCTCAAG  
25 GCTTTCTCTACCTGTGGATCTCACCTTGCCTTGGTCATTCTTTTCTATGGAGCAATCA  
CAGGGGTCTATATGAGCCCCTTATCCAATCACTCTACTGAAAAAGACTCAGCCGCAT  
CAGTCATTTTTATGGTTGTAGCACCTGTGTTGAATCCATTTCATTACAGTTTAAGAAA  
CAATGAACTGAAGGGGACTTTAAAAAAGACCCTAAGCCGACCGGGCGCGGTGGCTC  
ACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGGTGGATCATGAGGTCAGGA  
30 GATCGAGACCATCCTGGCTAACAAGGTGAAACCCCGT (SEQ ID NO.: 1)

MEGKNQTNISEFLLLGFSSWQQQVLLFALFLCLYLTGLFGNLLILLAIGSDHCLHTPMYF  
 FLANLSLVDLCLPSATVPKMLLNITQTQTISYPGCLAQMYFCMMFANMDNFLTVMMA  
 YDRYVAICHPLHYSTIMALRLCASLVAAPWVIAILNPLLHTLMMMAHLHFCSDNVIHHFFC  
 DINSLPLSCSDTSLNQLSVLATVGLIFVVPVCILVSYILIVSAVMKVPSAQGKLKAFSTC  
 5 GSHLALVILFYGAITGVYMSPLSNHSTEKDSAASVIFMVVAPVLNPFYSLRNNELKGTLK  
 KTLSPRGAVAHACNPSTLGGRGGWIMRSGDRDHPG (SEQ ID NO.: 2)

The NOV1 nucleic acid sequence has homology with several fragments of the human  
 olfactory receptor 17-93 (OLFR) (GenBank Accession No.: HSU76377), as shown in Table 3.  
 10 Also, the NOV1 polypeptide has homology (approximately 61% identity, 74% similarity) to  
 human olfactory receptor, family 1, subfamily F, member 8 (OLFR) (GenBank Accession No.:  
 XP007973), as is shown in Table 4. Furthermore, the NOV1 polypeptide has homology  
 (approximately 61% identity, 75% similarity) to a human olfactory protein (OLFR)(EMBL  
 Accession No.: 043749), as is shown in Table 5.

15 Overall amino acid sequence identity within the mammalian OR family ranges from 45%  
 to >80%. OR genes that are 80% or more identical to each other at the amino acid level are  
 considered by convention to belong to the same subfamily. See *Dryer and Berghard, Trends in  
 Pharmacological Sciences*, 1999, 20:413.

OR proteins have seven transmembrane  $\alpha$ -helices separated by three extracellular and  
 20 three cytoplasmic loops, with an extracellular amino-terminus and a cytoplasmic carboxy-  
 terminus. Multiple sequence alignment suggests that the ligand-binding domain of the ORs is  
 between the second and sixth transmembrane domains.

Thus, NOV1 is predicted to have a seven transmembrane region and is similar in that  
 region to a representative GPCR, *e.g.* dopamine (GPCR) (GenBank Accession No.: P20288), as  
 25 is shown in Table 6.

**TABLE 3**

30	NOV1:	1034	GGGCGCGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGGTGGATCAT	1093
	OLFR:	41200	GGATGCGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGTGGGCGGATCAT	41259
35	NOV1:	1094	GAGGTCAGGAGATCGAGACCATCCTGGCTAAC	1125 (SEQ ID NO. 33)
	OLFR:	41260	GAGGTCAGTTGTTTCGAGACCAACCTGGTCAAC	41291 (SEQ ID NO. 37)
	NOV1:	1032	CCGGGCGCGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGGTGGATC	1091
	OLFR:	1	CTGGGCTCGGTGGCTCACACGTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGGCGGATC	60



OLFR: 241 CGSHLAVVLLFYSTIIAVYFNPLSSHSAEKDTMATVLYTVVTPMLNPFIYSLRNRYLKG 299  
(SEQ ID NO. 30)

Where \* indicates identity and + indicates similarity.

5

**TABLE 6**

	NOV1:	48	AIGSDHCLHTPMYFFLANLSLVDLCLPSATVPKMLLNITQTQTISYPGCLAQMYFCMMF	107
	GPCR:	8	AVSREKALQTTTNYLIVSLAVADLLVATLVMPWVYLEVVGWKFRIHCDIFVTLDVMM	67
10	NOV1:	108	ANMDNFLTVMAYDRYVAICHPLHYSTIM-ALRLCASLVAAPWVIAILNPLLHTLMMHL	166
	GPCR:	68	CTASILNLCAISIDRYTAVAMPMLYNTRYSSKRRVTVMIAIVWVLSFTISCPMLFGLNNT	127
	NOV1:	167	HFCSDNVIHHFFCDINSLPLSCSDTSLNQLSVLATVGLIFVVPVSVILVSYILIVSAVM	226
	GPCR:	128	DQN-----ECIIANPAFVVYSSIVS--FYVPPFIVTLLVYIKIYIVLR	167
15	NOV1:	227	KVPSAQGKLK 236 (SEQ ID NO. 31)	
	GPCR:	168	RRRKRVTNKR 177 (SEQ ID NO. 32)	

Because the OR family of the GPCR superfamily is a group of related proteins specifically located at the ciliated surface of olfactory sensory neurons in the nasal epithelium and are involved in the initial steps of the olfactory signal transduction cascade, NOV1 can be used to detect nasal epithelial neuronal tissue.

Based on its relatedness to the known members of the OR family of the GPCR superfamily, NOV1 satisfies a need in the art by providing new diagnostic or therapeutic compositions useful in the treatment of disorders associated with alterations in the expression of members of OR family-like proteins. Nucleic acids, polypeptides, antibodies, and other compositions of the present invention are useful in the treatment and/or diagnosis of a variety of diseases and pathologies, including by way of nonlimiting example, those involving neurogenesis, cancer and wound healing.

30

## NOV2

A NOV2 sequence according to the invention is a nucleic acid sequence encoding a polypeptide related to the human odorant receptor (OR) family of the G-protein coupled receptor (GPCR) superfamily of proteins. A NOV2 nucleic acid and its encoded polypeptide includes the sequences shown in Table 7. The disclosed nucleic acid (SEQ ID NO:3) is 1,040 nucleotides in length and contains an open reading frame (ORF) that begins with an ATG initiation codon at nucleotides 82-84 and ends with a TGA stop codon at nucleotides 1,012-1,014. The representative ORF encodes a 310 amino acid polypeptide (SEQ ID NO:4). Putative untranslated regions upstream and downstream of the coding sequence are underlined in SEQ ID NO: 3.

**TABLE 7.**

CCGAACAAGTTAAAATGAATCTGTTTTTAAACACTTCTCCTAAACCATGAGCATTAA  
 CTTGATTTCTCTGTCATAGGGATATGGGAGACAATATAACATCCATCAGAGAGTTC  
 CTCCTACTGGGATTTCCCGTTGGCCCAAGGATTCAGATGCTCCTCTTTGGGCTCTTCT  
 5 CCCTGTTCTACGTCTTCACCCTGCTGGGGAACGGGACCATACTGGGGCTCATCTCAC  
 TGGACTCCAGACTGCACGCCCCCATGTACTTCTTCCTCTCACACCTGGCGGTCTGTCG  
 ACATCGCCTACGCCTGCAACACGGTGCCCCGGATGCTGGTGAACCTCCTGCATCCAG  
 CCAAGCCCATCTCCTTTGCGGGCCGCATGATGCAGACCTTTCTGTTTTCCACTTTTGC  
 TGTCACAGAATGTCTCCTCCTGGTGGTGTATGTCCTATGATCTGTACGTGGCCATCTGC  
 10 CACCCCCTCCGATATTTGGCCATCATGACCTGGAGAGTCTGCATCACCTCGCGGTG  
 ACTTCCTGGACCACTGGAGTCCTTTTATCCTTGATTCATCTTGTGTTACTTCTACCTTT  
 ACCCTTCTGTAGGCCCCAGAAAATTTATCACTTTTTTTGTGAAATCTTGGCTGTTCTC  
 AAACCTGCCTGTGCAGATACCCACATCAATGAGAACATGGTCTTGGCCGGAGCAATT  
 TCTGGGCTGGTGGGACCCTTGTCCACAATTGTAGTTTCATATATGTGCATCCTCTGTG  
 15 CTATCCTTCAGATCCAATCAAGGGAAGTTCAGAGGAAAGCCTTCCGCACCTGCTTCT  
 CCCACCTCTGTGTGATTGGACTCGTTTATGGCACAGCCATTATCATGTATGTTGGACC  
 CAGATATGGGAACCCCAAGGAGCAGAAGAAATATCTCCTGCTGTTTCACAGCCTCTT  
 TAATCCCATGCTCAATCCCCTTATCTGTAGTCTTAGGAACTCAGAAGTGAAGAATAC  
 TTTGAAGAGAGTGCTGGGAGTAGAAAGGGCTTTATGAAAAGGATTATGGCATTGTG  
 20 ACTGACA (SEQ ID NO.: 3)

MGDNITSIREFLLLGFPVGPRIQMLLFGLFSLFYVFTLLGNGTILGLISLDSRLHAPMYFFLS  
 HLA VVDIAYACNTVPRMLVNLLHPAKPISFAGRMMQTFLFSTFAVTECLLLVMSYDLY  
 VAICHPLRYLAIMTWRVCITLAVTSWTTGVLLSLIHLVLLLPLPFCRPQKIYHFFCEILAVL  
 25 KLACADTHINENMVLGAISGLVGPLSTIVVSYMCILCAILQIQSREVQRKAFRTCFSHLC  
 VIGLVYGTAIMYVGPRYGNPKEQKKYLLLFHSLFNPMNLNPLICSLRNSEVKNTLKRVLG  
 VERAL (SEQ ID NO.: 4)

30 The NOV2 nucleic acid, polypeptide, antibodies and other compositions of the present  
 invention can be used to detect nasal epithelial neuronal tissue. A NOV2 nucleic acid was  
 identified on human chromosome 6.

The NOV2 nucleic acid sequence has a high degree of homology (99% identity) with a human genomic clone corresponding to chromosome 6 (CHR6) (GenBank Accession No.: AL135904), as shown in Table 8. Additionally, the NOV2 polypeptide has a high degree of homology (approximately 95% identity) to a human olfactory receptor (OLFR) (GenBank Accession No.: AL135904), as shown in Table 9. Furthermore, the NOV2 polypeptide has a high degree of homology (approximately 91% identity) to a human olfactory protein (OLFR) (EMBL Accession No.: AC005587), as shown in Table 10. Overall amino acid sequence identity within the mammalian OR family ranges from 45% to >80%. OR genes that are 80% or more identical to each other at the amino acid level are considered by convention to belong to the same subfamily. See *Dryer and Berghard, Trends in Pharmacological Sciences*, 1999, 20:413.

OR proteins have seven transmembrane  $\alpha$ -helices separated by three extracellular and three cytoplasmic loops, along with an extracellular amino-terminus and a cytoplasmic carboxy-terminus. Multiple sequence alignment suggests that the ligand-binding domain of the ORs is between the second and sixth transmembrane domains. Thus, NOV2 is predicted to have a seven transmembrane region, and is similar in that region to a representative GPCR, e.g. dopamine (GPCR) (GenBank Accession No.: P20288) as is shown in Table 11.

**TABLE 8**

20	NOV2: 1	ccgaacaagttaaaatgaatctgttttttaaacacttctcctaaaccatgagcattaactt	60
	CHR6: 22579	ccgaacaagttaaaatgaatctgttttttaaacacttctcctaaaccatgagcattaactt	22520
25	NOV2: 61	gatttcctctgtcatagggatatgggagacaataacatccatcagagagttcctccta	120
	CHR6: 22519	gatttcctctgtcatagggatatgggagacaataacatccatcagagagttcctccta	22460
30	NOV2: 121	ctgggatttcccggtggcccaaggattcagatgctcctctttgggctcttctcctgttc	180
	CHR6: 22459	ctgggatttcccggtggcccaaggattcagatgctcctctttgggctcttctcctgttc	22400
35	NOV2: 181	tacgtcttcaccctgctggggaacgggaccataactggggctcatctcactggactccaga	240
	CHR6: 22399	tacgtcttcaccctgctggggaacgggaccataactggggctcatctcactggactccaga	22340
40	NOV2: 241	ctgcacgcccccatgtacttcttctctcacacctggcggtcgtcgacatcgccctacgcc	300
	CHR6: 22339	ctgcacgcccccatgtacttcttctctcacacctggcggtcgtcgacatcgccctacgcc	22280
45	NOV2: 301	tgcaacacggtgccccggatgctggtgaacctcctgcatccagccaagcccatctccttt	360
	CHR6: 22279	tgcaacacggtgccccggatgctggtgaacctcctgcatccagccaagcccatctccttt	22220



5  
 NOV2: 361 gcgggcccgcacatgatgcagacctttctgttttccacttttgctgtcacagaatgtctcctc 420  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 CHR6: 22219 gcgggcccgcacatgatgcagacctttctgttttccacttttgctgtcacagaatgtctcctc 22160

10  
 NOV2: 421 ctggtggtgatgtcctatgatctgtacgtggccatctgccacccccctccgatatttggcc 480  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 CHR6: 22159 ctggtggtgatgtcctatgatctgtacgtggccatctgccacccccctccgatatttggcc 22100

15  
 NOV2: 481 atcatgacctggagagtctgcatcacccctcgcggtgacttcctggaccactggagtcctt 540  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 CHR6: 22099 atcatgacctggagagtctgcatcacccctcgcggtgacttcctggaccactggagtcctt 22040

20  
 NOV2: 541 ttatccttgattcatcttgtgttacttctacctttacccttctgtaggccccagaaaatt 600  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 CHR6: 22039 ttatccttgattcatcttgtgttacttctacctttacccttctgtaggccccagaaaatt 21980

25  
 NOV2: 601 tatcacnnnnnnngtgaaatcttggtgttctcaaacttgctgtgcagatacccacatc 660  
 ||||||| ||||||||||||||||||||||||||||||||||||||||||||  
 CHR6: 21979 tatcacttttttctgtgaaatcttggtgttctcaaacttgctgtgcagatacccacatc 21920

30  
 NOV2: 661 aatgagaacatggtccttgccggagcaatttctgggctggtgggacccttgccacaatt 720  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 CHR6: 21919 aatgagaacatggtccttgccggagcaatttctgggctggtgggacccttgccacaatt 21860

35  
 NOV2: 721 gtagtttcatatatgtgcatcctctgtgctatccttcagatccaatcaagggaagttcag 780  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 CHR6: 21859 gtagtttcatatatgtgcatcctctgtgctatccttcagatccaatcaagggaagttcag 21800

40  
 NOV2: 781 aggaaagccttccgcacctgcttctccacctctgtgtgattggactcgtttatggcaca 840  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 CHR6: 21799 aggaaagccttccgcacctgcttctccacctctgtgtgattggactcgtttatggcaca 21740

45  
 NOV2: 841 gccattatcatgtatggttgaccagatatgggaaccccaaggagcagaagaaatatctc 900  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 CHR6: 21739 gccattatcatgtatggttgaccagatatgggaaccccaaggagcagaagaaatatctc 21680

50  
 NOV2: 901 ctgctgtttcacagcctctttaatcccatgctcaatccccttatctgtagtcttaggaac 960  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 CHR6: 21679 ctgctgtttcacagcctctttaatcccatgctcaatccccttatctgtagtcttaggaac 21620

55  
 NOV2: 961 tcagaagtgaagaatactttgaagagagtgtgggagtagaaagggttttatgaaaagga 1020  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 CHR6: 21619 tcagaagtgaagaatactttgaagagagtgtgggagtagaaagggttttatgaaaagga 21560

60  
 NOV2: 1021 ttatggcattgtgactgaca 1040 (SEQ ID NO. 3)  
 ||||||||||||||||||  
 CHR6: 21559 ttatggcattgtgactgaca 21540 (SEQ ID NO. 34)

**TABLE 9**

NOV2: 51 DSRLHAPMYFFLSHLAVVDIAYACNTVPRMLVNLLHPAKPISFAGRMMQTFLFSTFAVTE 110  
 \*\*\*\*\*  
 OLFR: 13 DSRLHAPMYFFLSHLAVVDIAYACNTVPRMLVNLLHPAKPISFAGRMMQTFLFSTFAVTE 72

NOV2: 111 CLLLVVMSYDLYVAICHPLRYLAIMTWRVCITLAVTSWTTGVXXXXXXXXXXXXPFCRP 170  
 \*\*\*\*\*  
 OLFR: 73 CLLLVVMSYDLYVAICHPLRYLAIMTWRVCITLAVTSWTTGVLLSLIHLVLLLPLPFCRP 132

5 NOV2: 171 QKIYHFFCEILAVLKLACADTHINENMVLGAISGLVGPLSTIVVSYMCILCAILQIQSR 230  
 \*\*\*\*\*  
 OLFR: 133 QKIYHFFCEILAVLKLACADTHINENMVLGAISGLVGPLSTIVVSYMCILCAILQIQSR 192

10 NOV2: 231 EVQRKAFRTCFSHLCVIGLVYGTAIMYVGPRYGNPKEQKKYLLLFHSLFNPMLNPLICS 290  
 \*\*\*\*\*  
 OLFR: 193 EVQRKAFRTCFSHLCVIGLVYGTAIMYVGPRYGNPKEQKKYLLLFHSLFNPMLNPLICS 252

NOV2: 291 LRNSEVKNTLKRVLGVERAL 310 (SEQ ID NO. 35)  
 \*\*\*\*\*

15 OLFR: 253 LRNSEVKNTLKRVLGVERAL 272 (SEQ ID NO. 36)  
 Where \* indicates identity

**TABLE 10**

20 NOV2: 1 MGDNITSIREFLLLGFPVGPRIQMLLFGLFSLFYVFXXXXXXXXXXXXDSRLHAPMYF 60  
 \*\*\*\*\*  
 OLFR: 1 MGDNITSIREFLLLGFPVGPRIQMLLFGLFSLFYVFTLLGNGTILGLISLDSRLHAPMYF 60

25 NOV2: 61 FLSHLAVVDIAYACNTVPRMLVNLLHPAKPISFAGRMMQTFLFSTFAVTECLLLVMSYD 120  
 \*\*\*\*\*  
 OLFR: 61 FLSHLAVVDIAYACNTVPRMLVNLLHPAKPISFAGRMMQTFLFSTFAVTECLLLVMSYD 120

30 NOV2: 121 LYVAICHPLRYLAIMTWRVCITLAVTSWTTGVXXXXXXXXXXXXPFCRPQKIYHFFCEI 180  
 \*\*\*\*\*  
 OLFR: 121 LYVAICHPLRYLAIMTWRVCITLAVTSWTTGVLLSLIHLVLLLPLPFCRPQKIYHFFCEI 180

NOV2: 181 LAVLKLACADTHINENMVLGAISGLVGPLSTIVVSYMCILCAILQIQSREVQRKAFRTC 240  
 \*\*\*\*\*  
 OLFR: 181 LAVLKLACADTHINENMVLGAISGLVGPLSTIVVSYMCILCAILQIQSREVQRKAFRTC 240

35 NOV2: 241 FSHLCVIGLVYGTAIMYVGPRYGNPKEQKKYLLLFHSLFNPMLNPLICSLRNSEVKNTL 300  
 \*\*\*\*\*  
 OLFR: 241 FSHLCVIGLVYGTAIMYVGPRYGNPKEQKKYLLLFHSLFNPMLNPLICSLRNSEVKNTL 300

40 NOV2: 301 KRVLGVERAL 310 (SEQ ID NO. 4)  
 \*\*\*\*\*  
 OLFR: 301 KRVLGVERAL 310 (SEQ ID NO. 38)  
 Where \* indicates identity

**TABLE 11**

45 NOV2: 53 RLHAPMYFFLSHLAVVDIAYACNTVPRMLVNLLHPAKPISFAGRMMQTFLFSTFAVTECL 112  
 GPCR: 14 ALQTTTNYLIVSLAVADLLVATLVMPWVVYLEVVGEWKFSRIHCDIFVTLDVMMCTASIL 73

NOV2: 113 LLVMSYDLYVAICHPLRYLAIMTW-RVCITLAVTSWTTGVLLSLIHLVLLLPLPFCRPQ 171  
 GPCR: 74 NLCAISIDRYTAVAMPMLYNTRYSSKRRVTVMIAIVWVLSFTISCPMLFGLNNTDQNE-- 131

50 NOV2: 172 KIYHFFCEILAVLKLACADTHINENMVLGAISGLVGPLSTIVVSYMCILCAILQIQSRE 231  
 GPCR: 132 -CIIANPAF-----VYSSIVSFYVPFIVTLLVYIKIYIVLRRRRKRV 173

NOV2: 232 VQRK 235 (SEQ ID NO. 39)  
 GPCR: 174 NTKR 177 (SEQ ID NO. 40)

55



AGCCTGCACTGACACCCATGTGAAAGAGCTGGCTTTATTTAGCCTCAGCATCCTGGT  
AATTATGGTGCCTTTTCTGTTAATTCTCATATCCTATGGCTTCATAGTTAACACCATC  
CTGAAGATCCCCTCAGCTGAGGGCAAGAAGGCCTTTGTCACCTGTGCCTCACATCTC  
ACTGTGGTCTTTGTCCACTATGGCTGTGCCTCTATCATCTATCTGCGGCCCAAGTCCA  
5 AGTCTGCCTCAGACAAGGATCAGTTGGTGGCAGTGACCTACACAGTGGTTACTCCCT  
TACTTAATCCTCTTGTCTACAGTCTGAGGAACAAAGAGGTAAAACTGCATTGAAAA  
GAGTTCTTGGAATGCCTGTGGCAACCAAGATGAGCTAACAAAAATAATAATAAAA  
TTAACTAGGATAGTCACAGAAGAAATCAAAGGCATAAAATTTTCTGACCTTTAATGC  
ATGTCTCAGACAGTGTTTCCAAGGATTAAGACTACTCTTGCCTTTTTATTTTCTCC  
10 (SEQ ID NO.: 5)

MRGFNKTTVVTQFILVGFSSLGELQLLLFVIFLLLYLTILVANVTIMAVIRFSWTLHTPMY  
GFLFILSFSESCYTFVIIPQLLVHLLSDTKTISFMACATQLFFFLGFACTNCLLIAMGYDRY  
VAICHPLRYTLIINKRLGLELISLSGATGFFIALVATNLICDMRFCGPNRVNHYFCDMAPVI  
15 KLACTIONDTHVKELALFSLVIMVPFLLILISYGFIVNTILKIPSAEGKKAFVTCASHLTVVF  
VHYGCASIIYLRPKSKSASDKDQLVAVTYTVVTPLLNPLVYSLRNKEVKTALKRVLGMP  
VATKMS (SEQ ID NO.: 6)

The OR family of the GPCR superfamily is a group of related proteins specifically  
20 located at the ciliated surface of olfactory sensory neurons in the nasal epithelium and are  
involved in the initial steps of the olfactory signal transduction cascade. Accordingly, the NOV3  
nucleic acid, polypeptide, antibodies and other compositions of the present invention can be used  
to detect nasal epithelial neuronal tissue. A NOV3 nucleic acid was identified on human  
chromosome 1.

25 The NOV3 nucleic acid sequence has a high degree of homology (99% identity) with a  
human genomic clone corresponding to chromosome 1 (CHR1) (GenBank Accession  
No.:AL121986), as is shown in Table 13. Also, the NOV3 polypeptide has homology  
(approximately 50% identity, 70% similarity) to a human olfactory receptor (OLFR) (GenBank  
Accession No.: F20722), as is shown in Table 14. Overall amino acid sequence identity within  
30 the mammalian OR family ranges from 45% to >80%. OR genes that are 80% or more identical  
to each other at the amino acid level are considered by convention to belong to the same  
subfamily See *Dryer and Berghard, Trends in Pharmacological Sciences*, 1999, 20:413. OR

proteins have seven transmembrane  $\alpha$ -helices separated by three extracellular and three cytoplasmic loops, with an extracellular amino-terminus and a cytoplasmic carboxy-terminus. Multiple sequence alignment suggests that the ligand-binding domain of the ORs is between the second and sixth transmembrane domains. NOV3 is predicted to have a seven transmembrane region, and is similar in that region to a representative GPCR, *e.g.* dopamine (GPCR) (GenBank Accession No.: P20288) as is shown in Table 15.

**TABLE 13**

10	NOV3: 1	aagaagttcttcagatgcgaggtttcaacaaaaccactgtgggttacacagttcatcctgg	60
	CHR1: 145895	aagaagttcttcagatgcgaggtttcaacaaaaccactgtgggttacacagttcatcctgg	145836
15	NOV3: 61	tgggtttctccagcctgggggagctccagctgctactttttgtcatctttcttctcctat	120
	CHR1: 145835	tgggtttctccagcctgggggagctccagctgctactttttgtcatctttcttctcctat	145776
20	NOV3: 121	acttgacaatcctggtggccaatgtgaccatcatggccgttattcgcttcagctggactc	180
	CHR1: 145775	acttgacaatcctggtggccaatgtgaccatcatggccgttattcgcttcagctggactc	145716
25	NOV3: 181	tccacactcccatgtatggctttctattcatcctttcattttctgagtcctgctacactt	240
	CHR1: 145715	tccacactcccatgtatggctttctattcatcctttcattttctgagtcctgctacactt	145656
30	NOV3: 241	ttgtcatcatccctcagctgctgggtccacctgctctcagacaccaagaccatctcctca	300
	CHR1: 145655	ttgtcatcatccctcagctgctgggtccacctgctctcagacaccaagaccatctcctca	145596
35	NOV3: 301	tggcctgtgccacccagctgttctttttccttggctttgcttgcaccaactgcctcctca	360
	CHR1: 145595	tggcctgtgccacccagctgttctttttccttggctttgcttgcaccaactgcctcctca	145536
40	NOV3: 361	ttgctgtgatgggatatgatcgctatgtagcaatttgtcaccctctgaggtacacactca	420
	CHR1: 145535	ttgctgtgatgggatatgatcgctatgtagcaatttgtcaccctctgaggtacacactca	145476
45	NOV3: 421	tcataaataaaaaggctgggggttgaggttgatttctctctcaggggccacaggtttcttta	480
	CHR1: 145475	tcataaataaaaaggctgggggttgaggttgatttctctctcaggggccacaggtttcttta	145416
50	NOV3: 481	ttgctttggtggccaccaacctcatttgtgacatgcgtttttgtggccccaacaggggta	540
	CHR1: 145415	ttgctttggtggccaccaacctcatttgtgacatgcgtttttgtggccccaacaggggta	145356
55	NOV3: 541	accactatcttctgtgacatggcacctgttatcaagtttagcctgcactgacacccatgtga	600
	CHR1: 145355	accactatcttctgtgacatggcacctgttatcaagtttagcctgcactgacacccatgtga	145296
60	NOV3: 601	aagagctggcctttatcttagcctcagcatcctggtaattatgggtgccttttctgttaattc	660
	CHR1: 145295	aagagctggcctttatcttagcctcagcatcctggtaattatgggtgccttttctgttaattc	145236





**TABLE 15**

	NOV3 :	43	NVTIMAVIRFSWTLHTPMYGFLFILSFSESCYTFVVIIPQLLVHLLSDTKTISFMACATQL	102
	GPCR :	2	NVLVCMASVSREKALQTTTNYLIVSLAVADLLVATLVMPWVYLEVVGWKFSTRHCDIFV	61
5	NOV3 :	103	FFFLGFACTNCLLIAMGYDRYVAICHPLRYTLIIN-KRLGLELISLSGATGFFIALVAT	161
	GPCR :	62	TLDVMMCTASILNLCAISIDRYTAVAMPMLYNTRYSSKRRVTVMIAIVWVLSFTISCPML	121
	NOV3 :	162	NLICDMRFCGPNRVNHYFCDMAPVIKLACTDTHVKELALFSLVIMVPFLLILISYGF	221
10	GPCR :	122	FGLNNTDQNEC-----IIANPAFVVYSSIVSFYVPFIVTLLVYIK	161
	NOV3 :	222	IVNTILKI	229 (SEQ ID NO. 45)
	GPCR :	162	IYIVLRRR	169 (SEQ ID NO. 46)

The OR family of the GPCR superfamily is a group of related proteins specifically located at the ciliated surface of olfactory sensory neurons in the nasal epithelium and are involved in the initial steps of the olfactory signal transduction cascade. Accordingly, in one embodiment, the NOV3 nucleic acid, polypeptide, antibodies and other compositions of the present invention can be used to detect nasal epithelial neuronal tissue.

Based on its relatedness to the known members of the OR family of the GPCR superfamily, NOV3 satisfies a need in the art by providing new diagnostic or therapeutic compositions useful in the treatment of disorders associated with alterations in the expression of members of OR family-like proteins. Nucleic acids, polypeptides, antibodies, and other compositions of the present invention are useful in the treatment and/or diagnosis of a variety of diseases and pathologies, including by way of nonlimiting example, those involving neurogenesis, cancer and wound healing.

#### NOV4

A NOV4 sequence according to the invention is a nucleic acid sequence encoding a polypeptide related to the human odorant receptor (OR) family of the G-protein coupled receptor (GPCR) superfamily of proteins. The NOV3 nucleic acid sequence (SEQ ID NO.: 5) was further analyzed by exon linking and the resulting sequence was identified as NOV4. A NOV4 nucleic acid and its encoded polypeptide includes the sequences shown in Table 16. The disclosed nucleic acid (SEQ ID NO:7) is 1,090 nucleotides in length and contains an open reading frame (ORF) that begins with an ATG initiation codon at nucleotides 15-17 and ends with a TAA stop codon at nucleotides 1,061-1,063. The representative ORF encodes a 314 amino acid polypeptide (SEQ ID NO:8). Putative untranslated regions upstream and downstream of the coding sequence are underlined in SEQ ID NO: 7.



Overall amino acid sequence identity within the mammalian OR family ranges from 45% to >80%. OR genes that are 80% or more identical to each other at the amino acid level are considered by convention to belong to the same subfamily. See *Dryer and Berghard, Trends in Pharmacological Sciences*, 1999, 20:413. OR proteins have seven transmembrane  $\alpha$ -helices separated by three extracellular and three cytoplasmic loops, with an extracellular amino-terminus and a cytoplasmic carboxy-terminus. Multiple sequence alignment suggests that the ligand-binding domain of the ORs is between the second and sixth transmembrane domains. NOV4 is predicted to have a seven transmembrane region, and is similar in that region to a representative GPCR, *e.g.* dopamine (GPCR) (GenBank Accession No.: P20288) as is shown in Table 20.

**TABLE 17**

15	NOV4: 1	aagaagttcttcagatgcgagggttcaacaaaaccactgtgggttacacagttcatcctgg	60
	CHR1: 145895	aagaagttcttcagatgcgagggttcaacaaaaccactgtgggttacacagttcatcctgg	145836
20	NOV4: 61	tgggtttctccagcctgggggagctccagctgctactttttgtcatctttcttctcctat	120
	CHR1: 145835	tgggtttctccagcctgggggagctccagctgctgctttttgtcatctttcttctcctat	145776
25	NOV4: 121	acttgacaatcctgggtggccaatgtgaccatcatggccgttattcgcttcagctggactc	180
	CHR1: 145775	acttgacaatcctgggtggccaatgtgaccatcatggccgttattcgcttcagctggactc	145716
30	NOV4: 181	tccacactcccatgtatggctttctattcatcctttcattttctgagtcctgctacactt	240
	CHR1: 145715	tccacactcccatgtatggctttctattcatcctttcattttctgagtcctgctacactt	145656
35	NOV4: 241	ttgtcatcatcctcagctgctgggtccacctgctctcagacaccaagaccatctccctca	300
	CHR1: 145655	ttgtcatcatcctcagctgctgggtccacctgctctcagacaccaagaccatctccctca	145596
40	NOV4: 301	tggcctgtgccaccagctgttctttttccttggctttgcttgcaccaactgcctcctca	360
	CHR1: 145595	tggcctgtgccaccagctgttctttttccttggctttgcttgcaccaactgcctcctca	145536
45	NOV4: 361	ttgctgtgatgggatatgatcgctatgtagcaatttgtcacctctgaggtacacactca	420
	CHR1: 145535	ttgctgtgatgggatatgatcgctatgtagcaatttgtcacctctgaggtacacactca	145476
50	NOV4: 421	tcataaataaaaggctgggggttgagttgatttctctctcagggggccacaggtttcttta	480
	CHR1: 145475	tcataaataaaaggctgggggttgagttgatttctctctcagggggccacaggtttcttta	145416
55	NOV4: 481	ttgctttgggtggccaccaacctcatttgtgacatgcggtttttgtggccccaacaggggta	540
	CHR1: 145415	ttgctttgggtggccaccaacctcatttgtgacatgcggtttttgtggccccaacaggggta	145356

5 NOV4: 541 accactatctctgtgacatggcacctgttatcaagttagcctgcactgacacccatgtga 600  
 |||||  
 CHR1: 145355 accactatctctgtgacatggcacctgttatcaagttagcctgcactgacacccatgtga 145296

10 NOV4: 601 aagagctggctttatcttagcctcagcatcctggtaattatggcctttctgttaattc 660  
 |||||  
 CHR1: 145295 aagagctggctttatcttagcctcagcatcctggtaattatggcctttctgttaattc 145236

15 NOV4: 661 tcatatcctatggcttcatagtcacacccatcctgaagatccccctcagctgagggcaaga 720  
 |||||  
 CHR1: 145235 tcatatcctatggcttcatagtcacacccatcctgaagatccccctcagctgagggcaaga 145176

20 NOV4: 721 aggcctttgtcacctgtgcctcacatctcactgtggtccttgtccactatgactgtgcct 780  
 |||||  
 CHR1: 145175 aggcctttgtcacctgtgcctcacatctcactgtggtccttgtccactatgactgtgcct 145116

25 NOV4: 781 ctatcatctatctgcggcccaagtccaagtctgcctcagacaaggatcagttggtggcag 840  
 |||||  
 CHR1: 145115 ctatcatctatctgcggcccaagtccaagtctgcctcagacaaggatcagttggtggcag 145056

30 NOV4: 841 tgacctacgcagtgggttactcccttacttaatcctcttgtctacagtctgaggaacaaag 900  
 |||||  
 CHR1: 145055 tgacctacgcagtgggttactcccttacttaatcctcttgtctacagtctgaggaacaaag 144996

35 NOV4: 901 aggtaaaaactgcattgaaaagagttccttggaatgcctgtggcaaccaagatgagctaac 960  
 |||||  
 CHR1: 144995 aggtaaaaactgcattgaaaagagttccttggaatgcctgtggcaaccaagatgagctaac 144936

40 NOV4: 961 aaaaaataataataaaaattaactaggatagtcacagaagaaatcaaaggcataaaaatctt 1020  
 |||||  
 CHR1: 144935 aaaaaataataataaaaattaactaggatagtcacagaagaaatcaaaggcataaaaatctt 144876

45 NOV4: 1021 ctgacctttaatgcatgtctcagacagtgttccaaggattaagactactcttgcccttt 1080  
 |||||  
 CHR1: 144875 ctgacctttaatgcatgtctcagacagtgttccaaggattaagactactcttgcccttt 144816

50 NOV4: 1081 tatctctcc 1090 (SEQ ID NO. 4)  
 |||||  
 CHR1: 144815 tatctctcc 144806 (SEQ ID NO. 42)

**TABLE 18**

55 NOV4: 1 AAGAAGTTCTTCAGATGCGAGGTTTCAACAAAACCACTGTGGTTACACAGTTCATCCTGG 60  
 |||||  
 NOV3: 1 AAGAAGTTCTTCAGATGCGAGGTTTCAACAAAACCACTGTGGTTACACAGTTCATCCTGG 60

60 NOV4: 61 TGGGTTTCTCCAGCCTGGGGGAGCTCCAGCTGCTACTTTTGTTCATCTTTCTTCTCCTAT 120  
 |||||  
 NOV3: 61 TGGGTTTCTCCAGCCTGGGGGAGCTCCAGCTGCTGCTTTTGTTCATCTTTCTTCTCCTAT 120

65 NOV4: 121 ACTTGACAATCCTGGTGGCCAATGTGACCATCATGGCCGTTATTTCGCTTCAGCTGGACTC 180  
 |||||  
 NOV3: 121 ACTTGACAATCCTGGTGGCCAATGTGACCATCATGGCCGTTATTTCGCTTCAGCTGGACTC 180

NOV4: 181 TCCACACTCCCATGTATGGCTTTCTATTTCATCCTTTTCTGAGTCCTGCTACACTT 240  
 |||||  
 NOV3: 181 TCCACACTCCCATGTATGGCTTTCTATTTCATCCTTTTCTGAGTCCTGCTACACTT 240

NOV4: 241 TTGTCATCATCCCTCAGCTGCTGGTCCACCTGCTCTCAGACACCAAGACCATCTCCCTCA 300







## NOV5

A NOV5 sequence according to the invention is a nucleic acid sequence encoding a polypeptide related to the human odorant receptor (OR) family of the G-protein coupled receptor (GPCR) superfamily of proteins. A NOV5 nucleic acid and its encoded polypeptide includes the sequences shown in Table 21. The disclosed nucleic acid (SEQ ID NO:9) is 822 nucleotides in length and contains an open reading frame (ORF) that begins at nucleotide 6 and ends with a TGA stop codon at nucleotides 800-802. In addition, C indicates 'G' to 'C' substitutions in the sequence to correct stop codons. A representative ORF encodes a 265 amino acid polypeptide (SEQ ID NO:10). A putative untranslated region downstream of the coding sequence is underlined in SEQ ID NO: 9.

**TABLE 21**

CACACCCCATGTGCTTCTTCCTCTCCAAACTGTGCTCAGCTGACATCGGTTTTCACCT  
TGGCCATGGTTCCCAAGATGATTGTGAACATGCAGTCGCATAGCAGAGTCATCTCTT  
15 ATGAGGGCTGCCTGACACGGATGTCTTTCTTTGTCTTTTTGCATGTATGGAAGACAT  
GCTCCTGACTGTGATGGCCTATGACTGCTTTGTAGCCATCTGTGCGCCCTCTGCACTAC  
CCAGTCATCGTGAATCCTCACCTCTGTGTCTTCTTCGTCTTGGTGTCTTTTTCTTAG  
CCCGTTGGATTCCCAGCTGCACAGTTGGATTGTGTTACTATTACCATCATCAAGAA  
TGTGGAAATCACTAATTTTGTCTGTGAACCCTCTCAACTTCTCAACCTTGCTTGTTCT  
20 GACAGCGTCATCAATAACATATTCATATAATTCGATAGTACTATGTTTGGTTTCTTC  
CCATTTCAGGGATCCTTTTGTCTTACTATAAAATTGTCCCCTCCATTCTAAGGATGTC  
ATCGTCAGATGGGAAGTATAAAGGCTTCTCCACCTGTGGCTCTTACCTGGCAGTTGT  
TTGCTCATTTGATGGAACAGGCATTGGCATGTACCTGACTTCAGCTGTGTCACCACC  
CCCCAGGAATGGTGTGGTGGCGTCAGTGATGTATGCTGTGGTCACCCCCATGCTGAA  
25 CCTTTTCATCTCAGCCTAGGAAAGAGGGATATACAAAGTGTCTGCGGAGGCTGTGC  
AGCAGAACAGTCGAATCTCATGATATGTTCCATCCTTTTTCTTGTGTGGGTGAGAAA  
GGGCAACCACATTAAA (SEQ ID NO.: 9)

PMCFFLSKLCSADIGFTLAMVPMKIVNMQSHSRVISYEGCLTRMSFFVLFACMEDMLLT  
30 VMAYDCFVAICRPLHYPVIVNPHLCVFFVLVSFFLSPLDSQLHSWIVLLFTIKNVEITNFV  
CEPSQLNLACSDSVINNIFIYFDSTMFGFLPISGILLSYYKIVPSILRMSSSDGKYKGFSTC  
GSYLAVVCSFDGTGIGMYLTSVSPPPRNGVVASVMYAVVTPMLNLFYSLGKRDIQSV  
LRRLCSRTVESHD MFHPFSCVG (SEQ ID NO.: 10)

35 The OR family of the GPCR superfamily is a group of related proteins specifically located at the ciliated surface of olfactory sensory neurons in the nasal epithelium and are involved in the initial steps of the olfactory signal transduction cascade. Accordingly, the NOV5 nucleic acid, polypeptide, antibodies and other compositions of the present invention can be used to detect nasal epithelial neuronal tissue.

The NOV5 nucleic acid sequence has a high degree of homology (94% identity) with a human genomic clone containing an OR pseudogene (OLFR) (GenBank Accession No.:AF065864), as is shown in Table 22. The NOV5 polypeptide has homology (approximately 67% identity, 79% similarity) to a human olfactory receptor (OLFR) (EMBL Accession No.:043789), as is shown in Table 23. Overall amino acid sequence identity within the mammalian OR family ranges from 45% to >80%. OR genes that are 80% or more identical to each other at the amino acid level are considered by convention to belong to the same subfamily. See *Dryer and Berghard, Trends in Pharmacological Sciences*, 1999, 20:413. OR proteins have seven transmembrane  $\alpha$ -helices separated by three extracellular and three cytoplasmic loops, with an extracellular amino-terminus and a cytoplasmic carboxy-terminus. Multiple sequence alignment suggests that the ligand-binding domain of the ORs is between the second and sixth transmembrane domains. NOV5 is predicted to have a seven transmembrane region, and is similar in that region to a representative GPCR, *e.g.* dopamine (GPCR) (GenBank Accession No.: P20288) as is shown in Table 24.

**TABLE 22**

NOV5:	1	CACACCCCCATGTGCTTCTTCCTCTCCAACTGTGCTCAGCTGACATCGGTTTCACCTTG	60
OLFR:	136	CACACCCCCATGTGCTTCTTCCTCTCCAACTGTGCTGGGCTGACATCGGTTTCACCTTG	195
NOV5:	61	GCCATGGTTCCCAAGATGATTGTGAACATGCAGTCGCATAGCAGAGTCATCTCTTATGAG	120
OLFR:	196	GCCACGGTTCCTAAGATGATTGTGGACATGCAGTCTCATACCAGAGTCATCTCTTATGAG	255
NOV5:	121	GGCTGCCTGACACGGATGTCTTTCTTTGTCCTTTTTCATGTATGGAAGACATGCTCCTG	180
OLFR:	256	GGCTGCCTGACACGGATATCTTTCTTGGTCTTTTTCATGTATAGAAGACATGCTCCTG	315
NOV5:	181	ACTGTGATGGCCTATGACTGCTTTGTAGCCATCTGTGCGCCCTCTGCACTACCCAGTCATC	240
OLFR:	316	ACTGTGATGGCCTATGACTGCTTTGTAGCCATCTGTGCGCCCTCTGCACTACCCAGTCATC	375
NOV5:	241	GTGAATCCTCACCTCTGTGTCTTCTTCGTCTTGGTGTCTTTTTCCTTAGCCCGTTGGAT	300
OLFR:	376	GTGAATCCTCACCTCTGTGTCTTCTTCCTTTTGGTATACTTTTTCCTTAGCTTGGTGGAT	435
NOV5:	301	TCCCAGCTGCACAGTTGGATTGTGTTACTATTACCATCATCAAGAATGTGGAAATCACT	360
OLFR:	436	TCCCAGCTGCACAGTTGGATTGTGTTACAATTACCATCATCAAGAATGTGGAAATCTCT	495
NOV5:	361	AATTTTGTCTGTGAACCCTCTCAACTTCTCAACCTTGCTTGTTCTGACAGCGTCATCAAT	420
OLFR:	496	AATTTTGTCTGTGACCCCTCTCAACTTCTCAAACCTGCCTGTTCTGACAGCGTCATCAAT	555
NOV5:	421	AACATATTCATATATTTTCGATAGTACTATGTTTGGTTTTCTTCCCATTTCAGGGATCCTT	480
OLFR:	556	AGCATATTCATGTATTTCCATAGTACTATGTTTGGTTTTCTTCCCATTTCAGGGATCCTT	615
NOV5:	481	TTGTCTTACTATAAAATTGTCCCCTCCATTCTAAGGATGTCATCGTCAGATGGGAAGTAT	540



acid, polypeptide, antibodies and other compositions of the present invention can be used to detect nasal epithelial neuronal tissue.

Based on its relatedness to the known members of the OR family of the GPCR superfamily, NOV5 satisfies a need in the art by providing new diagnostic or therapeutic compositions useful in the treatment of disorders associated with alterations in the expression of members of OR family-like proteins. Nucleic acids, polypeptides, antibodies, and other compositions of the present invention are useful in the treatment and/or diagnosis of a variety of diseases and pathologies, including by way of nonlimiting example, those involving neurogenesis, cancer and wound healing.

#### NOV6

A NOV6 sequence according to the invention is a nucleic acid sequence encoding a polypeptide related to the human odorant receptor (OR) family of the G-protein coupled receptor (GPCR) superfamily of proteins. A NOV6 nucleic acid and its encoded polypeptide includes the sequences shown in Table 25. The disclosed nucleic acid (SEQ ID NO:11) is 930 nucleotides in length and contains an open reading frame (ORF) that begins with an ATG initiation codon at nucleotides 22-24 and ends with a TAA stop codon at nucleotides 907-909. In addition, C indicates 'G' to 'C' substitutions in the sequence to correct stop codons. The representative ORF encodes a 294 amino acid polypeptide (SEQ ID NO:12). Putative untranslated regions up- and downstream of the coding sequence are underlined in SEQ ID NO: 11.

**TABLE 25.**

TTGCTGTCCCTGTCCCTGTCCATGTATATGGTCACGGTGCTGAGGAACCTGCT  
CAGCATCCTGGCTGTCAGCTCTGACTCCCCGCTCCACACCCCCATGTGCTTCT  
TCCTCTCCAAACTGTGCTCAGCTGACATCGGTTTCACCTTGGCCATGGTTCCC  
AAGATGATTGTGAACATGCAGTCGCATAGCAGAGTCATCTCTTATGAGGGCT  
GCCTGACACGGATGTCTTTCTTTGTCCTTTTTTGCATGTATGGAAGACATGCTC  
CTGACTGTGATGGCCTATGACTGCTTTGTAGCCATCTGTCGCCCTCTGCACTA  
CCCAGTCATCGTGAATCCTCACCTCTGTGTCTTCTTCGTCTTGGTGTCTTTTTT  
CCTTAGCCCGTTGGATTCCCAGCTGCACAGTTGGATTGTGTTACTATTACCA

TCATCAAGAATGTGGAAATCACTAATTTTGTCTGTGAACCCTCTCAACTTCTC  
AACCTTGCTTGTTCTGACAGCGTCATCAATAACATATTCATATATTTTCGATAG  
TACTATGTTTGGTTTTCTTCCCATTTCAGGGATCCTTTTGTCTTACTATAAAAT  
TGTCCCCTCCATTCTAAGGATGTCATCGTCAGATGGGAAGTATAAAGGCTTCT  
5 CCACCTGTGGCTCTTACCTGGCAGTTGTTTGCTCATTGATGGAACAGGCATT  
GGCATGTACCTGACTTCAGCTGTGTCACCACCCCCCAGGAATGGTGTGGTGG  
CGTCAGTGATGTATGCTGTGGTCACCCCCATGCTGAACCTTTTCATACTCAGC  
CTGGGAAAGAGGGATATACAAAGTGTCTGCGGAGGCTGTGCAGCAGAACA  
GTCGAATCTCATGATATGTTCCATCCTTTTTCTTGTGTGGGTGAGAAAGGGCA  
10 ACCACATTAAATCTCTACATCTGTAAATCCT (SEQ ID NO.: 11)

MYMVTVLRNLLSILAVSSDSPLHTPMCFFLSKLCSADIGFTLAMVPMIVNMQSHSRVIS  
YEGCLTRMSFFVLFACMEDMLLTVMAYDCFVAICRPLHYPVIVNPHLCVFFVLVSFFLSP  
LDSQLHSWIVLLFTIKNVEITNFVCEPSQLNLACSDSVINNIFIYFDSTMFGFLPISGILLSY  
15 YKIVPSILRMSSSDGKYKGFSTCGSYLAVVCSFDGTGIGMYLTSVSPPPRNGVASVMYA  
VVTPLNLFILSLGKRDIQSVLRRLCSRTVESHDMFHPFSCVGEKGQPH (SEQ ID NO.:  
12)

The OR family of the GPCR superfamily is a group of related proteins specifically  
located at the ciliated surface of olfactory sensory neurons in the nasal epithelium and are  
involved in the initial steps of the olfactory signal transduction cascade. Accordingly, the NOV6  
nucleic acid, polypeptide, antibodies and other compositions of the present invention can be used  
to detect nasal epithelial neuronal tissue.

The NOV6 nucleic acid sequence has a high degree of homology (94% identity) with a  
human genomic clone containing an OR pseudogene (OLFR) (GenBank Accession  
No.:AF065864), as is shown in Table 26. The NOV6 polypeptide has homology (approximately  
67% identity, 79% similarity) to a human olfactory receptor (OLFR) (EMBL Accession  
No.:043789), as is shown in Table 27. As shown in Table 28, the NOV6 polypeptide also has a  
high degree of homology (99% identity) with the NOV5 polypeptide. Overall amino acid  
sequence identity within the mammalian OR family ranges from 45% to >80%. OR genes that  
are 80% or more identical to each other at the amino acid level are considered by convention to  
belong to the same subfamily. See *Dryer and Berghard, Trends in Pharmacological Sciences*,  
1999, 20:413. Thus, NOV5 and NOV6 belong to the same OR subfamily.

OR proteins have seven transmembrane  $\alpha$ -helices separated by three extracellular and three cytoplasmic loops, with an extracellular amino-terminus and a cytoplasmic carboxy-terminus. Multiple sequence alignment suggests that the ligand-binding domain of the ORs is between the second and sixth transmembrane domains. NOV6 is predicted to have a seven transmembrane region, and is similar in that region to a representative GPCR, *e.g.* dopamine (GPCR) (GenBank Accession No.: P20288) as is shown in Table 29.

**TABLE 26**

10	NOV6: 10	ctgtccctgtccatgtatatgggtcacggtgctgaggaacctgctcagcatcctggctgtc	69
	OLFR: 58	ctgtccctgtccatgtatctgggtcacggtgctgaggaacctgctcatcatcctggctgtc	117
15	NOV6: 70	agctctgactccccgctccacacccccatgtgcttcttccctctccaaactgtgctcagct	129
	OLFR: 118	agctctgacccccacctccacacccccatgtgcttcttccctctccaaactgtgctgggct	177

NOV6: 130 gacatcgggtttcaccttggccatgggtcccaagatgattgtgaacatgcagtcgcatagc 189  
 |||||  
 OLFR: 178 gacatcgggtttcaccttggccacgggttcctaagatgattgtggacatgcagtcctcatacc 237

5

NOV6: 190 agagtcatctcttatgagggctgcctgacacggatgtctttctttgtcctttttgcatgt 249  
 |||||  
 OLFR: 238 agagtcatctcttatgagggctgcctgacacggatgtctttctttgtcctttttgcatgt 297

10

NOV6: 250 atggaagacatgctcctgactgtgatggcctatgactgctttgtagccatctgtcgccct 309  
 |||||  
 OLFR: 298 atagaagacatgctcctgactgtgatggcctatgactgctttgtagccatctgtcgccct 357

15

NOV6: 310 ctgcactaccagtcacgtgaatcctcacctctgtgtcttcttctgtctttggtgtccttt 369  
 |||||  
 OLFR: 358 ctgcactaccagtcacgtgaatcctcacctctgtgtcttcttctgtctttggtgtatacttt 417

20

NOV6: 370 ttccttagcccggttgattcccagctgcacagttggattgtgttactattcaccatcatc 429  
 |||||  
 OLFR: 418 ttccttagcttgttgattcccagctgcacagttggattgtgttacaattcaccatcatc 477

25

NOV6: 430 aagaatgtggaaatcactaattttgtctgtgaaccctctcaacttctcaaccttgcttgt 489  
 |||||  
 OLFR: 478 aagaatgtggaaatcctctaattttgtctgtgaaccctctcaacttctcaaccttgcttgt 537

30

NOV6: 490 tctgacagcgtcatcaataacatattcatatatttcgatagtactatgtttggttttctt 549  
 |||||  
 OLFR: 538 tctgacagcgtcatcaatagcatattcatgtatttccatagtactatgtttggttttctt 597

35

NOV6: 550 cccatttcagggatccttttgtcttactataaaaattgtccctccattctaaggatgtca 609  
 |||||  
 OLFR: 598 cccatttcagggatccttttgtcttactataaaaatcgtccctccattctaaggatttca 657

40

NOV6: 610 tcgtcagatgggaagtataaaggcttctccacctgtggtcttacctggcagttgtttgc 669  
 |||||  
 OLFR: 658 tcgtcagatgggaagtataaaggcttctccacctgtggtcttacctggcagttgtttgc 717

45

NOV6: 670 tcatttgatggaacaggcattggcatgtacctgacttcagctgtgtcaccacccccagg 729  
 |||||  
 OLFR: 718 tgattttatggaacaggcattggcgtgtacctgacttcagctgtgtcaccacccccagg 777

50

NOV6: 730 aatgggtgtggtggcgtcagtgatgtatgctgtggtcaccctcatgctgaaccttttcata 789  
 |||||  
 OLFR: 778 aatgggtgtggttagcgtcagtgatgtacgctgtggtcaccctcatgctgaaccttttcata 837

55

NOV6: 790 ctgagcctgggaagaggatatacaaagtgtcctgcggaggctgtgcagcagaacagtc 849  
 |||||  
 OLFR: 838 tacagcctgagaaacaggacatacaaagtgcctgcggaggctgtgcagcagaacagtc 897

60

NOV6: 850 gaatctcatgatatgttccatccttt 875 (SEQ ID NO. 59)  
 |||||  
 OLFR: 898 gaatctcatgatctgttccatccttt 923 (SEQ ID NO. 60)



## 5

[illegible]

## 25

25 NOV6: 25 PMCFFLSKLCSADIGFTLAMVPMIVNMQSHSRVISYEGCLTRMSFFVLFACMEDMLLTV 84  
 \*\*\*\*\*  
 NOV5: 1 PMCFFLSKLCSADIGFTLAMVPMIVNMQSHSRVISYEGCLTRMSFFVLFACMEDMLLTV 60  
  
 30 NOV6: 85 MAYDCFVAICRPLHYPVIVNPHLCXXXXXXXXXXXXXXXXXQLHSWIVLLFTTIKNVEITNF 144  
 \*\*\*\*\*  
 NOV5: 61 MAYDCFVAICRPLHYPVIVNPHLCVFFVLVSFFLSPLDSQLHSWIVLLFTTIKNVEITNF 120  
  
 NOV6: 145 VCEPSQLLNLACSDSVINNIFIYFDSTMFGFLPISGILLSYYKIVPSILRMSSSDGKYKG 204  
 \*\*\*\*\*  
 35 NOV5: 121 VCEPSQLLNLACSDSVINNIFIYFDSTMFGFLPISGILLSYYKIVPSILRMSSSDGKYKG 180  
  
 NOV6: 205 FSTCGSYLAVVCSFDGTGIGMYLTSVSPPPRNG-VASVMYAVVTPMLNLFILSLGKRDI 263  
 \*\*\*\*\*  
 NOV5: 181 FSTCGSYLAVVCSFDGTGIGMYLTSVSPPPRNGVVASVMYAVVTPMLNLFIIYSLGKRDI 240  
  
 40 NOV6: 264 QSVLRRLCSRTVESHDMPFSCVG 288 (SEQ ID NO. 63)  
 \*\*\*\*\*  
 NOV5: 241 QSVLRRLCSRTVESHDMPFSCVG 265 (SEQ ID NO. 10)  
 Where \* indicates identity.

## 45

<b><u>TABLE 29</u></b>				
50	NOV6:	9	NLLSILAVSSDSPLHTPMCFFLSKLCSADIGFTLAMVPKMIVNMQSHSRVISYEGCLTRM	68
	GPCR:	2	NVLVCMASVSREKALQTTTNYLIVSLAVADLLVATLVMPWVVYLEVVGWKFSSRIHCDIFV	61
	NOV6:	69	SFFVLFACMEDMLLTVMAYDCFVAICRPLHYPVIVNPHLCVFFVLVSFFLSPLDSQLHSW	128
	GPCR:	62	TLDVMMCTASILNLCAISIDRYTAVAMPMLYNTRYSSKRRV-----TVM	105
55	NOV6:	129	IVLLFTTIKNEITNFVCEPSQLNLACSDSVINNIFIYFDSTMFGFLPISGILLSYYKI	188
	GPCR:	106	IAIVWVLSFTISCPMLFG--LNNTDQNECIANPAFVVYSSIVSFYVPFIVTLLVYIKI	162
	NOV6:	189	VPSILRMSSS	198 (SEQ ID NO. 65)
	GPCR:	163	YIVLRRRRKR	172 (SEQ ID NO. 66)

The OR family of the GPCR superfamily is a group of related proteins specifically located at the ciliated surface of olfactory sensory neurons in the nasal epithelium and are involved in the initial steps of the olfactory signal transduction cascade. Accordingly, the NOV6 nucleic acid, polypeptide, antibodies and other compositions of the present invention can be used to detect nasal epithelial neuronal tissue.

Based on its relatedness to the known members of the OR family of the GPCR superfamily, NOV6 satisfies a need in the art by providing new diagnostic or therapeutic compositions useful in the treatment of disorders associated with alterations in the expression of members of OR family-like proteins. Nucleic acids, polypeptides, antibodies, and other compositions of the present invention are useful in the treatment and/or diagnosis of a variety of diseases and pathologies, including by way of nonlimiting example, those involving neurogenesis, cancer and wound healing.

#### NOV7

A NOV7 sequence according to the invention is a nucleic acid sequence encoding a polypeptide related to the human odorant receptor (OR) family of the G-protein coupled receptor (GPCR) superfamily of proteins. A NOV7 nucleic acid and its encoded polypeptide includes the sequences shown in Table 30. The disclosed nucleic acid (SEQ ID NO:13) is 930 nucleotides in length and contains an open reading frame (ORF) that begins with an ACG initiation codon at nucleotides 10-12 and ends with a TGA stop codon at nucleotides 882-884. In addition, C indicates 'G' to 'C' substitutions in the sequence to correct stop codons. The representative ORF encodes a 309 amino acid polypeptide (SEQ ID NO:12). Putative untranslated regions up- and downstream of the coding sequence are underlined in SEQ ID NO: 13

#### TABLE 30.

CACAGAGCCACGGAATCTCACAGGTGTCTCAGAAATTCCTCCTCCTGGGACTCTCAGA  
GGATCCAGAACTGCAGCCGGTCCTCGCTTTGCTGTCCCTGTCCCTGTCCATGTATCTG  
GTCACAGTGCTGAGGAACCTGCTCAGCATCCCGGCTGTCAGCTCTGACTCCCACCTC  
CACACCCCCACGTACTTCTTCCTCTCCATCCTGTGCTGGGCTGACATCGGTTTCACCT  
CGGCCACGGTTCCCAAGATGATTGTGGACATGCAGTGGTATAGCAGAGTCATCTCTC  
ATGCGGGCTGCCTGACACAGATGTCTTTCTTGGTCCTTTTTTGCATGTATAGAAGGCAT  
GCTCCTGACTGTAATGGCCTATGACTGCTTTGTAGGCATCTATCGCCCTCTGCACTAC  
CCAGTCATCGTGAATCCTCATCTCTGTGTCTTCTTTGTTTTGGTGTCTTTTTCTCTTAG  
CCTGTTGGATTCCCAGCTGCACAGTTGGATTGTGTTACAATTACCATCATCAAGAA  
TGTGGAAATCTCTAATTTTGTCTGTGACCCCTCTCAACTTCTCAAACCTTGCCTCTTAT  
GACAGCGTCATCAATAGCATATTCATATATTTTCGATAGTACAATGTTTGGTTTTCTTC

CTATTTTCAGGGATCCTTTTCATCTTACTATAAAAATTGTCCCCTCCATTCTAAGGATGTC  
 ATCGTCAGATGGGAAGTATAAACTTTCTCCACCTATGGCTCTCACCTAGCATTGTGTT  
 TGCTCATTATTTATGGAACAGGCATTGACATGTACCTGGCTTCAGCTATGTCACCAACC  
 CCCAGGAATGGTGTGGTGGTGTGTCAGTGATGTAAGCTGTGGTCACCCCCATGCTGAAC  
 5 CTTTTCATCTACAGCCTGAGAAACAGGGACATACAAAGTGCCCTGCGGAGGCTGCG  
 CAGCAGAAC (SEQ ID NO.: 13)

TEPRNLTGVSEFLLLGLSEDPELQPVLLLSLSLSMYLVTVLRNLLSIPAVSSDSHLHTPTY  
 FFLSILCWADIGFTSATVPKMIVDMQWYSRVISHAGCLTQMSFLVLFACIEGMLLTVMAY  
 10 DCFVGIYRPLHYPVIVNPHLCVFFVLVSFFLSLLDSQLHSWIVLQFTIKNVEISNFVCDPSQ  
 LLKLASYDSVINSIFIYFDSTMFGFLPISGILSSYYKIVPSILRMSSSDGKYKTFSTYGSHLAF  
 VCSFYGTGIDMYLASAMSPTRNGVVVSVMXAVVTPMLNLFYSLRNRDIQSALRRLRS  
 R (SEQ ID NO.: 14)

15

The OR family of the GPCR superfamily is a group of related proteins specifically located at the ciliated surface of olfactory sensory neurons in the nasal epithelium and are involved in the initial steps of the olfactory signal transduction cascade. The NOV7 nucleic acid, polypeptide, antibodies and other compositions of the present invention can be used to detect nasal epithelial neuronal tissue.

20

The NOV7 nucleic acid sequence has a high degree of homology (94% identity) with the human genomic clone pDJ392a17 from chromosome 11 (CHR11) (GenBank Accession No.:AC000385), as is shown in Table 31. The NOV7 polypeptide has homology (approximately 68% identity, 78% similarity) to a human olfactory receptor (OLFR) (EMBL Accession No.:043789), as is shown in Table 32.

25

Overall amino acid sequence identity within the mammalian OR family ranges from 45% to >80%. OR genes that are 80% or more identical to each other at the amino acid level are considered by convention to belong to the same subfamily. See *Dryer and Berghard, Trends in Pharmacological Sciences*, 1999, 20:413. OR proteins have seven transmembrane  $\alpha$ -helices separated by three extracellular and three cytoplasmic loops, with an extracellular amino-terminus and a cytoplasmic carboxy-terminus. Multiple sequence alignment suggests that the ligand-binding domain of the ORs is between the second and sixth transmembrane domains.

30

NOV7 is predicted to have a seven transmembrane region, and is similar in that region to a representative GPCR, e.g. dopamine (GPCR) (GenBank Accession No.: P20288) as is shown in Table 33.

35

### **TABLE 31.**

NOV7: 1 cacagagccacggaatctcacaggtgtctcagaattcctcctcctgggactctcagagga 60

5  
 NOV7: 61 tccagaactgcagccggtcctcgctttgctgtccctgtccctgtccatgtatctggtcac 120  
 CHR11: 126642 tccagaactgcagtcggtcctcgctttgctgtccctgtccctgtccctgaatctggtcac 126583  
 10  
 NOV7: 121 agtgctgaggaacctgctcagcatcccggtgtcagctctgactcccacctccacacccc 180  
 CHR11: 126582 ggtgctgaggaacctgctcagcatcctgggtgtcagctctgactccccctccacacccc 126523  
 15  
 NOV7: 181 cacgtacttcttctctccatcctgtgctgggctgacatcggtttcacctcggccacggt 240  
 CHR11: 126522 catgtacttcttctctccaacctgtgctgggctgacatcggtctcacctcggccacggt 126463  
 20  
 NOV7: 241 tcccaagatgattgtggacatgcagtggtatagcagagtcattctctcatgcgggctgcct 300  
 CHR11: 126462 tcccaaggatgattctggatatgcagtcgcatagcagagtcattctctcatgtgggctgcct 126403  
 25  
 NOV7: 301 gacacagatgtctttcttggtcctttttgcatgtatagaaggcatgctcctgactgtaat 360  
 CHR11: 126402 gacacagatgtctttcttggtcctttttgcatgtatagaaggcatgctcctgactgtgat 126343  
 30  
 NOV7: 361 ggccatgactgctttgtaggcatctatcgccctctgcaactaccagtcacgtgaatcc 420  
 CHR11: 126342 ggccatggctgctttgtagccatctgtcgccctctgcaactaccagtcacagtgaatcc 126283  
 35  
 NOV7: 421 tcattctctgtgtcttcttctgttttggtgtcctttttccttagcctgttgattcccagct 480  
 CHR11: 126282 tcacctctgtgtcttcttctgttttggtgtcctttttccttaacctgttgattcccagct 126223  
 40

NOV7: 481	gcacagttggattgtgttacaaattcaccatcatcaagaatgtggaaatctctaattttgt	540
CHR11: 126222	gcacagttggattgtgttacaattcaccatcatcaagaatgtggaaatctctaattttt	126163
NOV7: 541	ctgtgacccctctcaacttctcaaaacttgccctcttatgacagcgatcatcaatagcatatt	600
CHR11: 126162	ctgtgacccctctcagcttctcaaccttgccctgttctgacagcgatcatcaatagcatatt	126103
NOV7: 601	catatatttgcgatagtacaatgtttgggttttcttcctatttccagggatcctttcatctta	660
CHR11: 126102	catatatttgcgatagtactatgtttgggttttcttcccatctcagggatccttttgtctta	126043
NOV7: 661	ctataaaattgtccctccattctcaaggatgtcatcgatgggaagtataaaacttt	720
CHR11: 126042	ctataaaattgtccctccattctcaaggatgtcatcgatgggaagtataaagcctt	125983
NOV7: 721	ctccacctatggctctcacctagcatttgtttgctcattttatggaacaggcattgacat	780
CHR11: 125982	ctccacctatggctctcacctaggagttgtttgctgggttttatggaacagtcattggcat	125923
NOV7: 781	gtacctggcttcagctatgtcaccaacccccaggaatgggtgtggtggtgtcagtgatgta	840
CHR11: 125922	gtacctggcttcagccgtgtcaccacccccaggaatgggtgtggtggcatcagtgatgta	125863
NOV7: 841	agctgtgggtcacccccatgctgaaccttttcatctacagcctgagaaacaggacataca	900
CHR11: 125862	ggctgtgggtcacccccatgctgaaccttttcatctacagcctgagaaacaggacataca	125803
NOV7: 901	aagtgccttgcggaggctgcgcagcagaac	930 (SEQ ID NO. 13)
CHR11: 125802	aagtgccttgcggaggctgcgcagcagaac	125773 (SEQ ID NO. 68)

**TABLE 32**

NOV7:	179	PTYFFLSILCWADIGFTSATVPKMIQWYSRVISHAGCLTQMSFLVLFACIEGMLLTV	358
		* * * * *	
OLFR:	1	PMYFFLSNLSLADIGFTSTTVPKMIVDMQTHSRVISYEGCLTQMSFFVLFACMDDMLLSV	60
NOV7:	359	MAYDCFVGIIYRPLHYPIVINPHLCVFFVLVSFFLSLLDSQLHSHWIVLQFTTIKNVEISNF	538
		***** +*+** ** * +*+***+*****+ *+** * *+****	
OLFR:	61	MAYDRFVAICHPLHYRIIMNPRLCGFLILLSFFISLLDSQLHNLI MLQLTCFKDVIDISNF	120
NOV7:	539	VCDPSQLLKLASYSVINSIFIYFDSTMFGFLPISGILSSYYKIVPSILRMSSSDGKYKT	718
		***** * +* ** + *** +** ***** ***** ****+ +*****	
OLFR:	121	FCDPSQLHLRCSDTFINEMVIYFMGAIFGCLPISGILFSYYKIVSPILRVPTSDGKYKA	180
NOV7:	719	FSTYGSHLAFVCSFYGTGIDMYLASAMSPTPRNGVVSVSM*AVVTPMLNFLFIYSLRN RDI	898
		*** ***** ** *****+ **++**+ *+** +* *** ***** *****+**	
OLFR:	181	FSTCGSHLAVVCLFYGTGLVGYLSSAVLPSPRKSMVASVMYT VVTPMLNPFIYSLRNKDI	240
NOV7:	899	QSALRRRLRSR	928 (SEQ ID NO. 69)
		***** ** *	
OLFR:	241	QSALCRLHGR	250 (SEQ ID NO. 70)

Where \* indicates identity and + indicates similarity.

**TABLE 33**

NOV7: 44 NLLSIPAVSSDSHLHTPTYFFLSILCWADIGFTSATVPMIVDMQWYSRVISHAGCLTQM 103  
GPCR: 2 NVLVCMAVSREKALQTTTNYLIVSLAVADLLVATLVMPWVYLEVVGWKFSSRIHCDIFV 61

NOV7: 104 SFLVLFACIEGMLLTVMAYDCFVGIYRPLHYPVIVNPH 141 (SEQ ID NO. 71)

The OR family of the GPCR superfamily is a group of related proteins that are specifically located at the ciliated surface of olfactory sensory neurons in the nasal epithelium and are involved in the initial steps of the olfactory signal transduction cascade. Accordingly, the NOV7 nucleic acid, polypeptide, antibodies and other compositions of the present invention can be used to detect nasal epithelial neuronal tissue.

Based on its relatedness to the known members of the OR family of the GPCR superfamily, NOV7 satisfies a need in the art by providing new diagnostic or therapeutic compositions useful in the treatment of disorders associated with alterations in the expression of members of OR family-like proteins. Nucleic acids, polypeptides, antibodies, and other compositions of the present invention are useful in the treatment and/or diagnosis of a variety of diseases and pathologies, including by way of nonlimiting example, those involving neurogenesis, cancer and wound healing.

#### NOV8

A NOV8 sequence according to the invention is a nucleic acid sequence encoding a polypeptide related to the human odorant receptor (OR) family of the G-protein coupled receptor (GPCR) superfamily of proteins. A NOV8 nucleic acid and its encoded polypeptide includes the sequences shown in Table 34. The disclosed nucleic acid (SEQ ID NO:15) is 994 nucleotides in length and contains an open reading frame (ORF) that begins with an ATG initiation codon at nucleotides 27-29 and ends with a TGA stop codon at nucleotides 969-971. The representative ORF encodes a 314 amino acid polypeptide (SEQ ID NO:16). Putative untranslated regions up- and downstream of the coding sequence are underlined in SEQ ID NO: 15.

**TABLE 34.**

TGCAGCTAAAGTGCATTGTGTAAAACATGGGGGATGTGAATCAGTCGGTGGCCTCA  
GACTTCATTCTGGTGGGCCTCTTCAGTCACTCAGGATCACGCCAGCTCCTCTTCTCCC  
TGGTGGCTGTCATGTTTGTTCATAGGCCTTCTGGGCAACACCGTTCTTCTTCTTGAT  
CCGTGTGGACTCCCGGCTCCATACACCCATGTACTTCCTGCTCAGCCAGCTCTCCCTG  
TTTGACATTGGCTGTCCCATGGTCAACATCCCCAAGATGGCATCAGACTTTCTGCGG  
GGAGAAGGTGCCACCTCCTATGGAGGTGGTGCAGCTCAAATATTCTTCCTCACACTG  
ATGGGTGTGGCTGAGGGCGTCCTGTTGGTCCTCATGTCTTATGACCGTTATGTTGCTG  
TGTGCCAGCCCCTGCAGTATCCTGTACTTATGAGACGCCAGGTATGTCTGCTGATGA  
TGGGCTCCTCCTGGGTGGTAGGTGTGCTCAACGCCTCCATCCAGACCTCCATCACCC  
TGCATTTTCCCTACTGTGCCTCCCGTATTGTGGATCACTTCTTCTGTGAGGTGCCAGC  
CCTACTGAAGCTCTCCTGTGCAGATACCTGTGCCTACGAGATGGCGCTGTCCACCTC





OLFR: 1 MGTDNQTWVSEFILLGLSSDWDTRVSLFVLFLVMYVVTVLGNCLIVLLIRLDSRLHTPMY 60

NOV8: 61 FLLSQLSLFDIGCPMVTIPKMASDFLRGEGATSYGGGAAQIFFLTLMGVAEGVLLVLMYSY 120  
 \* \*+ \*\*\* \*+ +++++ + \*\* \* + \*\*\*\*+\* +\* \* \*\*\* +\*\*

5 OLFR: 61 FFLTNLSSLVDVSYATSVVPQLLAHFLAEHKAIPFQSCAAQLFFSLALGGIEFVLLAVMAY 120

NOV8: 121 DRYVAVCQPLQYPVLMRRQVCLLMMGSSWVVGVLNASIQTSITLHFPYCASRIVDHHFFCE 180  
 \*\*\*\*\* \*\* \*+ \*+ + +\*\*\*\* \* +++ \*\*\*\*+\* \* \* ++ \*\*\* \*\*

10 OLFR: 121 DRYVAVCDALRYSAIMHGGLCARLAITSWVSGFISSPVQTAITFQLPMCRNKFIDHISCE 180

NOV8: 181 VPALLKLSCADTCAYEMALSTSGVLILMLPLSLIATSYGHVLQAVLSMRSEEARHKAVTT 240  
 + \*\*\*\*+\* \* \* + \*+ + \* \*\*\*\*+\* \* \*+ \*\* ++ +\* +++ \* \* \* \*

OLFR: 181 LLAVVRLACVDTSSNEVTIMVSSIVLLMTPLCLVLLSYIQIISTILKIQSREGRKKAFHT 240

15 NOV8: 241 CSSHITVVGLFYGA AVFMYMPCAYHSPQQDNVVSFLFYSLVTPTLNPLIYSLRNPEVWMA 300  
 \*+\*\*\*\*+\* \* \* \*+\* \*+ \* + \* \*+ + \*+\*\*\*\*+\*+\* \*\*\*\*+\*\*\*\*\* \* \* \*

OLFR: 241 CASHLTVVALCYGVAIIFTYIQPHSSPSVLQEKLFSVFYAILTPMLNPMIYSLRNKEVKGA 300

NOV8: 301 LVKVL 305 (SEQ ID NO. 73)  
 \*\*

20 OLFR: 301 WQKLL 305 (SEQ ID NO. 74)  
 Where \* indicates identity and + indicates similarity.

**TABLE 36**

NOV8: 27 MGDVNQSVASDFILVGLFSHSGSRQLLFSVLVAVMFVIGLLGNTVLLFLIRVDSRLHTPMY 206  
 \*\* \*\*+ \*\*\*\*+\* \* +\* \*\* \* \*\*++ +\*\*\* +++ \*\*\*\*+\*\*\*\*\*

30 OLFR: 1 MGTDNQTWVSEFILLGLSSDWDTRVSLFVLFLVMYVVTVLGNCLIVLLIRLDSRLHTPMY 60

NOV8: 207 FLLSQLSLFDIGCPMVTIPKMASDFLRGEGATSYGGGAAQIFFLTLMGVAEGVLLVLMYSY 386  
 \* \*+ \*\*\* \*+ +++++ + \*\* \* + \*\*\*\*+\* +\* \* \*\*\* +\*\*

OLFR: 61 FFLTNLSSLVDVSYATSVVPQLLAHFLAEHKAIPFQSCAAQLFFSLALGGIEFVLLAVMAY 120

35 NOV8: 387 DRYVAVCQPLQYPVLMRRQVCLLMMGSSWVVGVLNASIQTSITLHFPYCASRIVDHHFFCE 566  
 \*\*\*\*\* \*\* \*+ \*+ + +\*\*\*\* \* +++ \*\*\*\*+\* \* \* ++ \*\*\* \*\*

OLFR: 121 DRYVAVCDALRYSAIMHGGLCARLAITSWVSGFISSPVQTAITFQLPMCRNKFIDHISCE 180

NOV8: 567 VPALLKLSCADTCAYEMALSTSGVLILMLPLSLIATSYGHVLQAVLSMRSEEARHKAVTT 746  
 + \*\*\*\*+\* \* \* + \*+ + \* \*\*\*\*+\* \* \*+ \*\* ++ +\* +++ \* \* \* \*

40 OLFR: 181 LLAVVRLACVDTSSNEVTIMVSSIVLLMTPLCLVLLSYIQIISTILKIQSREGRKKAFHT 240

NOV8: 747 CSSHITVVGLFYGA AVFMYMPCAYHSPQQDNVVSFLFYSLVTPTLNPLIYSLRNPEVWMA 926  
 \*+\*\*\*\*+\* \* \* \*+\* \*+ \* + \* \*+ + \*+\*\*\*\*+\*+\* \*\*\*\*+\*\*\*\*\* \* \* \*

45 OLFR: 241 CASHLTVVALCYGVAIIFTYIQPHSSPSVLQEKLFSVFYAILTPMLNPMIYSLRNKEVKGA 300

NOV8: 927 LVKVLSR-AGL 956 (SEQ ID NO. 75)  
 \*\* + \*\*

50 OLFR: 301 WQKLLWKFSGL 311 (SEQ ID NO. 76)  
 Where \* indicates identity and + indicates similarity.

**TABLE 37**

NOV8: 41 GNTVLLFLIRVDSRLHTPMYFLLSQLSLFDIGCPMVTIPKMASDFLRGEGATSYGGGAAQ 100

55 GPCR: 1 GNVLCMAVSREKALQTTTNYLIVSLAVADLLVATLVMPWVYLEVVGWKFRIHCDIF 60

NOV8: 101 IFFLTLMGVAEGVLLVLMYSYDRYVAVCQPLQYPVLM-RRQVCLLMMGSSWVVGVLNASIQ 159

GPCR: 61 VTLDVMMCTASILNLCAISIDRYTAVAMPMLYNTYSSKRRVTVMIAIVVLSFTISCPM 120

60 NOV8: 160 TSITLHFPYCASRIVDHHFFCEVPALLKLSCADTCAYEMALSTSGVLILMLPLSLIATSYG 219

GPCR: 121 LFGLNNTDQN-----ECIIA--NPAFVVYSSIVSFYVPFIVTLLVYI 160

NOV8: 220 HVLQAVLSMRSEEA 233 (SEQ ID NO. 77)

The OR family of the GPCR superfamily is a group of related proteins located at the ciliated surface of olfactory sensory neurons in the nasal epithelium. The OR family is involved in the initial steps of the olfactory signal transduction cascade. Accordingly, the NOV8 nucleic acid, polypeptide, antibodies and other compositions of the present invention can be used to detect nasal epithelial neuronal tissue.

Based on its relatedness to the known members of the OR family of the GPCR superfamily, NOV8 satisfies a need in the art by providing new diagnostic or therapeutic compositions useful in the treatment of disorders associated with alterations in the expression of members of OR family-like proteins. Nucleic acids, polypeptides, antibodies, and other compositions of the present invention are useful in the treatment and/or diagnosis of a variety of diseases and pathologies, including by way of nonlimiting example, those involving neurogenesis, cancer and wound healing.

#### NOV9

A NOV9 sequence according to the invention is a nucleic acid sequence encoding a polypeptide related to the human odorant receptor (OR) family of the G-protein coupled receptor (GPCR) superfamily of proteins. A NOV9 nucleic acid and its encoded polypeptide includes the sequences shown in Table 38. The NOV8 nucleic acid sequence (SEQ ID NO.: 15) was further analyzed by exon linking, and the resulting sequence was identified as NOV9. The disclosed nucleic acid (SEQ ID NO:17) is 994 nucleotides in length and contains an open reading frame (ORF) that begins with an ATG initiation codon at nucleotides 28-30 and ends with a TAG stop codon at nucleotides 979-981. The representative ORF encodes a 317 amino acid polypeptide (SEQ ID NO:18). Putative untranslated regions up- and downstream of the coding sequence are underlined in SEQ ID NO: 17.

**TABLE 38.**

TGCAGCTAAAGTGCATTGTGTAAACTATGGGGGATGTGAATCAGTCGGTGGCCTC  
 AGACTTCATTCTGGTGGGCCTCTTCAGTCACTCAGGATCACGCCAGCTCCTCTTCTCC  
 CTGGTGGCTGTCATGTTTGTTCATAGGCCTTCTGGGCAACACCGTTCTTCTTCTTGA  
 TCCGTGTGGACTCCCGGCTCCACACACCCATGTACTTCCTGCTCAGCCAGCTCTCCCT  
 GTTTGACATTGGCTGTCCCATGGTCAACATCCCCAAGATGGCATCAGACTTTCTGCG  
 GGGAGAAGGTGCCACCTCCTATGGAGGTGGTGCAGCTCAAATATTCTTCCTCACACT  
 GATGGGTGTGGCTGAGGGCGTCCTGTTGGTCCTCATGTCTTATGACCGTTATGTTGCT  
 GTGTGCCAGCCCCTGCAGTATCCTGTACTTATGAGACGCCAGGTATGTCTGCTGATG

ATGGGCTCCTCCTGGGTGGTAGGTGTGCTCAACGCCTCCATCCAGACCTCCATCACC  
 CTGCATTTTCCCTACTGTGCCTCCCGTATTGTGGATCACTTCTTCTGTGAGGTGCCAG  
 CCCTACTGAAGCTCTCCTGTGCAGATACCTGTGCCTACGAGATGGCGCTGTCCACCT  
 CAGGGGTGCTGATCCTAATGCTCCCTCTTTCCTCATCGCCACCTCCTACGGCCACGT  
 5 GTTGCAGGCTGTTCTAAGCATGCGCTCAGAGGAGGCCAGACACAAGGCTGTCACCA  
 CCTGCTCCTCGCACATCACGGTAGTGGGGCTCTTTTATGGTGCCGCCGTGTTTCATGTA  
 CATGGTGCCTTGCGCCTACCACAGTCCACAGCAGGATAACGTGGTTTCCCTCTTCTA  
 TAGCCTTGTCACCCCTACACTCAACCCCTTATCTACAGTCTGAGGAATCCGGAGGT  
 GTGGATGGCTTTGGTCAAAGTGCTTAGCAGAGCTGGACTCAGGCAAATGTGCATGAC  
 10 TACATAGAACTGCTGGTGAGA (SEQ ID NO.: 17)

MGDVNQSVASDFILVGLFSHSGSRQLLFSLVAVMFVIGLLGNTVLLFLIRVDSRLHTPMY  
 FLLSQLSLFDIGCPMVTIPKMASDFLRGEGATSYGGGAAQIFFLTLMGVAEGVLLVLMSY  
 DRYVAVCQPLQYPVLMRRQVCLLMGSSWVVGVLNASIQTSITLHFPYCASRIVDHFFC  
 15 EVPALLKLSCADTCAYEMALSTSGVLILMLPLSLIATSYGHVLQAVLSMRSEEARHKAVT  
 TCSSHITVVGLFYGAAVFMYMVPCAYHSPQQDNVVSFLFYSLVTPTLNPLIYSLRNPEVW  
 MALVKVLSRAGLRQMCMTT (SEQ ID NO.: 18)

The OR family of the GPCR superfamily is a group of related proteins specifically  
 20 located at the ciliated surface of olfactory sensory neurons in the nasal epithelium that are  
 involved in the initial steps of the olfactory signal transduction cascade. Accordingly, the NOV9  
 nucleic acid, polypeptide, antibodies and other compositions of the present invention can be used  
 to detect nasal epithelial neuronal tissue.

The NOV9 polypeptide has homology (approximately 44% identity, 65% similarity) to  
 25 the human olfactory receptor family 2 subfamily F, member 1 (OLFR) (EMBL Accession  
 No.:NP 036501), as is shown in Table 39. The NOV9 polypeptide also has a high degree of  
 homology (99% identity) to the NOV8 polypeptide as shown in Table 40. Overall amino acid  
 sequence identity within the mammalian OR family ranges from 45% to >80%. OR genes that  
 are 80% or more identical to each other at the amino acid level are considered by convention to  
 30 belong to the same subfamily. See *Dryer and Berghard, Trends in Pharmacological Sciences*,  
 1999, 20:413. Thus NOV8 and NOV9 belong to the same subfamily of ORs.

OR proteins have seven transmembrane  $\alpha$ -helices separated by three extracellular and  
 three cytoplasmic loops, with an extracellular amino-terminus and a cytoplasmic carboxy-  
 terminus. Multiple sequence alignment suggests that the ligand-binding domain of the ORs is  
 35 between the second and sixth transmembrane domains. NOV9 is predicted to have a seven  
 transmembrane region, and is similar in that region to a representative GPCR, *e.g.* dopamine  
 (GPCR) (GenBank Accession No.: P20288) as is shown in Table 41.

**TABLE 39**

5	NOV9: 1	MGDVNQSVASDFILVGLFSHSGSRQLLFSLVAVMFVIGLLGNTVLLFLIRVDSRLHTPMY 60	
	OLFR: 1	MGTDNQTWVSEFILLGLSSDWDTRVSLFVLFLVMYVVTVLGNCLIVLLIRLDSRLHTPMY 60	
10	NOV9: 61	FLLSQLSLFDIGCPMVTIPKMASDFLRGEGATSYGGGAAQIFFLTLMGVAEGVLLVLMSY 120	
	OLFR: 61	FFLTNLSSLVDVSYATSVVPQLLAHFLAEHKAIPFQSCAAQLFFSLALGGIEFVLLAVMAY 120	
15	NOV9: 121	DRYVAVCQPLQYPVLMRRQVCLLMGSSWVVGVLNASIQTSITLHFPYCASRIVDHFFCE 180	
	OLFR: 121	DRYVAVCDALRYSAIMHGGLCARLAITSWVSGFISSPVQTAITFQLPMCRNKFIDHISCE 180	
20	NOV9: 181	VPALLKLSCADTCAYEMALSTSGVLILMLPLSLIATSYGHVLQAVLSMRSEEARKHKAHTT 240	
	OLFR: 181	LLAVVRLACVDTSNEVTIMVSSIVLLMTPLCLVLLSYIQIISTILKIQSREGRKKAFHT 240	
25	NOV9: 241	CSSHITVVGLFYGAAVFMYMVPCAYHSPQQDNVVSFLYSLVTPTLNPLIYSLRNPEVWMA 300	
	OLFR: 241	CASHLTVVALCYGVAIFTYIQPHSSPSVLQEKLFVYFVAILTPMLNPMIYSLRNKEVKGA 300	
30	NOV9: 301	LVKVL 305 (SEQ ID NO. 79)	
	OLFR: 301	WQKLL 305 (SEQ ID NO. 80)	
Where * indicates identity and + indicates similarity.			

**TABLE 40**

30	NOV9: 1	MGDVNQSVASDFILVGLFSHSGSRQLLFSLVAVMFVIGLLGNTVLLFLIRVDSRLHTPMY 60	
	NOV8: 1	MGDVNQSVASDFILVGLFSHSGSRQLLFSLVAVMFVIGLLGNTVLLFLIRVDSRLHTPMY 60	
35	NOV9: 61	FLLSQLSLFDIGCPMVTIPKMASDFLRGEGATSYGGGAAQIFFLTLMGVAEGVLLVLMSY 120	
	NOV8: 61	FLLSQLSLFDIGCPMVTIPKMASDFLRGEGATSYGGGAAQIFFLTLMGVAEGVLLVLMSY 120	
40	NOV9: 121	DRYVAVCQPLQYPVLMRRQVCLLMGSSWVVGVLNASIQTSITLHFPYCASRIVDHFFCE 180	
	NOV8: 121	DRYVAVCQPLQYPVLMRRQVCLLMGSSWVVGVLNASIQTSITLHFPYCASRIVDHFFCE 180	
45	NOV9: 181	VPALLKLSCADTCAYEMALSTSGVLILMLPLSLIATSYGHVLQAVLSMRSEEARKHKAHTT 240	
	NOV8: 181	VPALLKLSCADTCAYEMALSTSGVLILMLPLSLIATSYGHVLQAVLSMRSEEARKHKAHTT 240	
50	NOV9: 241	CSSHITVVGLFYGAAVFMYMVPCAYHSPQQDNVVSFLYSLVTPTLNPLIYSLRNPEVWMA 300	
	NOV8: 241	CSSHITVVGLFYGAAVFMYMVPCAYHSPQQDNVVSFLYSLVTPTLNPLIYSLRNPEVWMA 300	
55	NOV9: 301	LVKVLSRAGLRQMCMTT 317 (SEQ ID NO. 17)	
	NOV8: 301	LVKVLSRAGLRQMC--- 314 (SEQ ID NO. 15)	
Where * indicates identity.			

**TABLE 41**

60	NOV9: 41	GNTVLLFLIRVDSRLHTPMYFLLSQLSLFDIGCPMVTIPKMASDFLRGEGATSYGGGAAQ 100	
	GPCR: 1	GNVLVCMASVREKALQTTNYLIVSLAVADLLVATLVMPWVYLEVVGWKFVSRHCDIF 60	
65	NOV9: 101	IFFLTLMGVAEGVLLVLMSYDRYVAVCQPLQYPVLM-RRQVCLLMGSSWVVGVLNASIQ 159	
	GPCR: 61	VTLDVMMCTASILNLCAISIDRYTAVAMPMLYNTRYSSKRRVTVMIAIVWVLSFTISCPM 120	

NOV9: 160 TSITLHFPYCASRIVDHFFCEVPALLKLSCADTCAYEMALSTSGVLILMLPLSLIATSYG 219  
GPCR: 121 LFGLNNTDQN-----ECIIA--NPAFVVYSSIVSFYVPFIVTLLVYI 160

5 NOV9: 220 HVLQAVLSMRSEEA 233 (SEQ ID NO. 83)  
GPCR: 161 KIYIVLRRRRKRNV 174 (SEQ ID NO. 84)

The OR family of the GPCR superfamily is a group of related proteins specifically located at the ciliated surface of olfactory sensory neurons in the nasal epithelium that are involved in the initial steps of the olfactory signal transduction cascade. Accordingly, the NOV9 nucleic acid, polypeptide, antibodies and other compositions of the present invention can be used to detect nasal epithelial neuronal tissue.

Based on its relatedness to the known members of the OR family of the GPCR superfamily, NOV9 satisfies a need in the art by providing new diagnostic or therapeutic compositions useful in the treatment of disorders associated with alterations in the expression of members of OR family-like proteins. Nucleic acids, polypeptides, antibodies, and other compositions of the present invention are useful in the treatment and/or diagnosis of a variety of diseases and pathologies, including by way of nonlimiting example, those involving neurogenesis, cancer and wound healing.

#### NOV10

A NOV10 sequence according to the invention is a nucleic acid sequence encoding a polypeptide related to the human odorant receptor (OR) family of the G-protein coupled receptor (GPCR) superfamily of proteins. A NOV10 nucleic acid and its encoded polypeptide includes the sequences shown in Table 42. The disclosed nucleic acid (SEQ ID NO:19) is 1,077 nucleotides in length and contains an open reading frame (ORF) that begins with an ATG initiation codon at nucleotides 31-33 and ends with a TAG stop codon at nucleotides 1,030-1,032. The representative ORF encodes a 318 amino acid polypeptide (SEQ ID NO:20). Putative untranslated regions up- and downstream of the coding sequence are underlined in SEQ ID NO: 19. Exon linking was used to confirm the sequence.

**TABLE 42.**

CAGGTTTCATTGACAAGGTCATACCAACCAGATGAATCCAGCAAATCATTCCCAGGT  
GGCAGGATTTGTTCTACTGGGGCTCTCTCAGGTTTGGGAGCTTCGGTTTGTCTTCTTC  
ACTGTTTTCTCTGCTGTGTATTTTATGACTGTAGTGGGAAACCTTCTTATTGTGGTCA  
TAGTGACCTCCGACCCACACCTGCACACAACCATGTATTTTCTCTTGGGCAATCTTTC

TTTCCTGGACTTTTGGCTACTCTTCCATCACAGCACCTAGGATGCTGGTTGACTTGCTC  
 TCAGGCAACCCTACCATTTCCTTTGGTGGATGCCTGACTCAACTCTTCTTCTTCCACT  
 TCATTGGAGGCATCAAGATCTTCCTGCTGACTGTCATGGCGTATGACCGCTACATTG  
 CCATTTCCCAGCCCCTGCACTACACGCTCATTATGAATCAGACTGTCTGTGCACTCCT  
 5 TATGGCAGCCTCCTGGGTGGGGGGCTTCATCCACTCCATAGTACAGATTGCATTGAC  
 TATCCAGCTGCCATTCTGTGGGCCTGACAAGCTGGACAACCTTTTATTGTGATGTGCCT  
 CAGCTGATCAAATTGGCCTGCACAGATAACCTTTGTCTTAGAGCTTTTAATGGTGTCTA  
 ACAATGGCCTGGTGACCCTGATGTGTTTTCTGGTGCTTCTGGGATCGTACACAGCAC  
 TGCTAGTCATGCTCCGAAGCCACTCACGGGAGGGCCGCAGCAAGGCCCTGTCTACCT  
 10 GTGCCTCTCACATTGCTGTGGTGACCTTAATCTTTGTGCCTTGCATCTACGTCTATAC  
 AAGGCCTTTTTCGGACATTCCCCATGGACAAGGCCGTCTCTGTGCTATACACAATTGT  
 CACCCCCATGCTGAATCCTGCCATCTATACCCTGAGAAACAAGGAAGTGATCATGGC  
 CATGAAGAAGCTGTGGAGGAGGAAAAAGGACCCTATTGGTCCCCTGGAGCACAGAC  
 CCTTACATTAGCAGAGGCAGTGACCTGAGAATCTGAAAGATGCTACAGGGTATTAG  
 15 CAGAGGCAGTGACCTGAGAATCTGAAAGATGCTACAGGGTATTAG (SEQ ID NO.:19)

MNPANHSQVAGFVLLGLSQVWELRFVFFTVFSAVYFMTVVGNLLIVVIVTSDPHLHTTM  
 YFLLGNLSFLDFCYSSITAPRMLVDLLSGNPTISFGGCLTQLFFFHFIGGIKIFLLTVMAYDR  
 YIAISQPLHYTLIMNQTVCALLMAASWVGGFHHSIVQIALTIQLPFCGPKLDNFYCDVPQ  
 20 LIKLACTDTFVLELLMVSNNGLVTLMCFLVLLGSYTALLVMLRSHSREGRSKALSTCASH  
 IAVVTLIFVPCIYVYTRPFRTPMDKA VSVLYTIVTPMLNPAIYTLRNKEVIMAMKKLWR  
 RKKDPIGPLEHRPLH (SEQ ID NO.: 20)

25 The OR family of the GPCR superfamily is a group of related proteins specifically located at the ciliated surface of olfactory sensory neurons in the nasal epithelium and are involved in the initial steps of the olfactory signal transduction cascade. Accordingly, the NOV10 nucleic acid, polypeptide, antibodies and other compositions of the present invention can be used to detect nasal epithelial neuronal tissue.

30 The NOV10 polypeptide has homology (approximately 55% identity, 72% similarity) to the olfactory receptor MOR83 (OLFR) (EMBL Accession No.:BAA86125), as is shown in Table 43. Overall amino acid sequence identity within the mammalian OR family ranges from 45% to >80%. OR genes that are 80% or more identical to each other at the amino acid level are considered by convention to belong to the same subfamily. See *Dryer and Berghard, Trends in*  
 35 Pharmacological Sciences, 1999, **20**:413. OR proteins have seven transmembrane  $\alpha$ -helices separated by three extracellular and three cytoplasmic loops, with an extracellular amino-terminus and a cytoplasmic carboxy-terminus. Multiple sequence alignment suggests that the ligand-binding domain of the ORs is between the second and sixth transmembrane domains. NOV10 is predicted to have a seven transmembrane region, and is similar in that region to a



representative GPCR, *e.g.* dopamine (GPCR) (GenBank Accession No.: P20288) as is shown in Table 44.

**TABLE 43**

5	NOV10:	79	MNPANHSQVAGFVLLGLSQVWELRFVFFTVFSVAVYFMTVVGNLLIVVIVTSDPHLHTTMY	258
			* * ++* *+ ****+ * * +** *+ * +++++* ****+ * * * *	
	OLFR:	1	MGALNQTRVTEFIFLGLTDNWNVLEILFFVPFTVTYMLTLLGNFLIVVTIVFTPRLHNPMY	60
10	NOV10:	259	FLLGNLSFLDFCYSSITAPRMLVDLLSGNPTISFGGCLTQLFFFHFIGGIKIFLLTVMAY	438
			* * ****+* +++++* +*** * * **** *+ **** * ++++++****	
	OLFR:	61	FFLSNLSFIDICHSSVTVPKMLEGLLLERKTISFDNCIAQLFFLHLFACSEIFLLTIMAY	120
15	NOV10:	439	DRYIAISQPLHYTLIMNQTVCALLMAASWVGFIHSIVQIALTIQLPFCGPDKLDNFYCD	618
			****+** ****+ +** * * *+ * ****+ ****+ ****+****+ +++++**	
	OLFR:	121	DRYVAICIPHYSNVMNMKVCVQLVFALWLGGTIHSIVQTLTIRLPYCGPNIIDSIFCD	180
20	NOV10:	619	VPQLIKLACTDTFVLELLMVSNNGLVTLMCFLVLLGSYTALLVMLRSHSREGRSKALSTC	798
			** +*****++ +++++*+ +++++*+ *+ ****+* * * * * ****	
	OLFR:	181	VPPVIKLACTDTYLTGILIVSNSGTISLVCFLALVTSYTVILFSLRKKSAEGRRKALSTC	240
25	NOV10:	799	ASHIAVVTILFVPCIVVYTRPFRTPMDKAVSVLYTIVTPMLNPAIYTLRNKEVIMAMKK	978
			++* **** * ****+***** +* +** **** ****+***** ****+*** **	
	OLFR:	241	SAHFMVVTLLFFGPCIFLYTRPDSSFSIDKVVSVFYTVVTPLLNPLIYTLRNEEVKTAMKH	300
	NOV10:	979	LWRRK 993 (SEQ ID NO. 85)	
			* ++*	
	OLFR:	301	LRQRR 305 (SEQ ID NO. 86)	

Where \* indicates identity and + indicates similarity.

**TABLE 44**

30	NOV10:	41	GNLLIVVIVTSDPHLHTTMYFLLGNLSFLDFCYSSITAPRMLVDLLSGNPTISFGGCLTQ	100
	GPCR:	1	GNVLVCMASVREKALQTTTNYLIVSLAVADLLVATLVMPWVYLEVVGWKFRIHCDIF	60
35	NOV10:	101	LFFFHFIGGIKIFLLTVMAYDRYIAISQPLHYTLIMNQ-TVCALLMAASWVGFIHSIVQ	159
	GPCR:	61	VTLDVMMCTASILNLCAISIDRYTAVAMPMLYNTRYSSKRRVTVMIAIVWVLSFTISCPM	120
	NOV10:	160	IALTIQLPFCGPDKLDNFYCDVPQLIKLACTDTFVLELLMVSNNGLVTLMCFLVLLGSYT	219
	GPCR:	121	LFGLNNTDQNE-----CIIANPAFVVY--SSIVSFYVPFIVTLLVYI	160
40	NOV10:	220	ALLVMLRSHSREGRSKA 236 (SEQ ID NO. 87)	
	GPCR:	161	KIYIVLRRRRKRVTNR 177 (SEQ ID NO. 88)	

The OR family of the GPCR superfamily is a group of related proteins specifically located at the ciliated surface of olfactory sensory neurons in the nasal epithelium that are involved in the initial steps of the olfactory signal transduction cascade. Accordingly, the NOV10 nucleic acid, polypeptide, antibodies and other compositions of the present invention can be used to detect nasal epithelial neuronal tissue.

Based on its relatedness to the known members of the OR family of the GPCR superfamily, NOV10 satisfies a need in the art by providing new diagnostic or therapeutic compositions useful in the treatment of disorders associated with alterations in the expression of members of OR family-like proteins. Nucleic acids, polypeptides, antibodies, and other



compositions of the present invention are useful in the treatment and/or diagnosis of a variety of diseases and pathologies, including by way of nonlimiting example, those involving neurogenesis, cancer and wound healing.

## 5 NOV11

A NOV11 sequence according to the invention is a nucleic acid sequence encoding a polypeptide related to the human odorant receptor (OR) family of the G-protein coupled receptor (GPCR) superfamily of proteins. A NOV11 nucleic acid was discovered by exon linking analysis of NOV2 (SEQ ID NO.: 3). A NOV11 nucleic acid and its encoded polypeptide includes the  
 10 sequences shown in Table 45. The disclosed nucleic acid (SEQ ID NO:21) is 1,012 nucleotides in length and contains an open reading frame (ORF) that begins with an ATG initiation codon at nucleotides 54-56 and ends with a TGA stop codon at nucleotides 984-986. The representative ORF encodes a 310 amino acid polypeptide (SEQ ID NO:22). Putative untranslated regions up- and downstream of the coding sequence are underlined in SEQ ID NO: 21.

15 **TABLE 45.**

AAACACTTCTCCTAAACCATGAGCATTAAGTTGATTTCCTCTGTCATAGGGATATGG  
GAGACAATATAACATCCATCACAGAGTTCTCCTACTGGGATTTCCCGTTGGCCCAA  
GGATTCAGATGCTCCTCTTTGGGCTCTTCTCCCTGTTCTACGTCTTCACCCTGCTGGG  
 20 GAACGGGACCATACTGGGGCTCATCTCACTGGACTCCAGACTGCACGCCCCCATGTGA  
CTTCTTCCTCTCACACCTGGCGGTGTCGACATCGCCTACGCCTGCAACACGGTGCC  
CCGGATGCTGGTGAACCTCCTGCATCCAGCCAAGCCCATCTCCTTTGCGGGCCGCAT  
GATGCAGACCTTTCTGTTTTCCACTTTTGCTGTACAGAAATGTCTCCTCCTGGTGGTG  
ATGTCCTATGATCTGTACGTGGCCATCTGCCACCCCCTCCGATAATTGGCCATCATGA  
 25 CCTGGAGAGTCTGCATCACCTCGCGGTGACTTCCTGGACCACTGGAGTCCTTTTAT  
CCTTGATTCATCTTGTGTTACTTCTACCTTTACCCTTCTGTAGGCCCCAGAAAATTTA  
TCACTTTTTTTGTGAAATCTTGGCTGTTCTCAAAGTTGCCTGTGCAGATACCCACATC  
AATGAGAACATGGTCTTGGCCGGAGCAATTTCTGGGCTGGTGGGACCCTTGTCCACA  
ATTGTAGTTTCATATATGTGCATCCTCTGTGCTATCCTTCAGATCCAATCAAGGGAAG  
 30 TTCAGAGGAAAGCCTTCTGCACCTGCTTCTCCACCTCTGTGTGATTGGACTCTTTTA  
TGGCACAGCCATTATCATGTATGTTGGACCCAGATATGGGAACCCCAAGGAGCAGA  
AGAAATATCTCCTGCTGTTTCACAGCCTCTTTAATCCCATGCTCAATCCCCTTATCTG  
TAGTCTTAGGAACTCAGAAGTGAAGAATACTTTGAAGAGAGTGCTGGGAGTAGAAA  
GGGCTTTATGAAAAGGATTATGGCATTGTGACTGACA (SEQ ID NO.: 21)

35 MGDNITSITEFLLGFPVGPRIQMLLFGLFSLFYVFTLLGNGTILGLISLDSRLHAPMYFFLS  
 HLAVVDIAYACNTVPRMLVNLLHPAKPISFAGRMMQTFLFSTFAVTECLLLVMSYDLY  
 VAICHPLRYLAIMTWRCITLAVTSWTTGVLLSLIHLVLLLPLPFCRPQKIYHFFCEILAVL  
 KLACADTHINENMVLGAISGLVGPLSTIVVSVMCILCAILQIQSREVQRKAFCTCFSHLC  
 40 VIGLFYGTAIMYVGPRYGNPKEQKKYLLLFHSLFNPMNLPLICSLRNSEVKNTLKRVLG  
 VERAL (SEQ ID NO.: 22)



NOV11: 301 KRVLGVERAL 310 (SEQ ID NO.: 64)  
 \*\*\*\*\*  
 OLFR: 301 KRVLGVERAL 310 (SEQ ID NO.: 67)

5 Where \* indicates identity.

**TABLE 47**

10 NOV11: 1 MGDNITSITEFLLLGFPVGPRIQMLLFGLFSLFYVFTLLGNGTILGLISLDSRLHAPMYF 60  
 \*\*\*\*\*  
 NOV2: 1 MGDNITSIREFLLLGFPVGPRIQMLLFGLFSLFYVFTLLGNGTILGLISLDSRLHAPMYF 60  
 NOV11: 61 FLSHLAVVDIAYACNTVPRMLVNLLHPAKPISFAGRMMQTFLFSTFAVTECLLVVMSYD 120  
 \*\*\*\*\*  
 15 NOV2: 61 FLSHLAVVDIAYACNTVPRMLVNLLHPAKPISFAGRMMQTFLFSTFAVTECLLVVMSYD 120  
 NOV11: 121 LYVAICHPLRYLAIMTWRVCITLAVTSWTTGVLLSLIHLVLLPLPFCRPQKIYHFFCEI 180  
 \*\*\*\*\*  
 NOV2: 121 LYVAICHPLRYLAIMTWRVCITLAVTSWTTGVLLSLIHLVLLPLPFCRPQKIYHFFCEI 180  
 20 NOV11: 181 LAVLKLACADTHINENMVLAGAISGLVGPLSTIVVSYMCILCAILQIQSREVQRKAFCTC 240  
 \*\*\*\*\*  
 NOV2: 181 LAVLKLACADTHINENMVLAGAISGLVGPLSTIVVSYMCILCAILQIQSREVQRKAFRTC 240  
 25 NOV11: 241 FSHLCVIGLFYGTAIMYVGPRYGNPKEQKKYLLLFHSLFNPMLNPLICSLRNSEVKNTL 300  
 \*\*\*\*\*  
 NOV2: 241 FSHLCVIGLVYGTAIMYVGPRYGNPKEQKKYLLLFHSLFNPMLNPLICSLRNSEVKNTL 300  
 NOV11: 301 KRVLGVERAL 310 (SEQ ID NO.: 22)  
 \*\*\*\*\*  
 30 NOV2: 301 KRVLGVERAL 310 (SEQ ID NO.: 4)  
 Where \* indicates identity.

**TABLE 48**

35 NOV11: 53 RLHAPMYFFLSHLAVVDIAYACNTVPRMLVNLLHPAKPISFAGRMMQTFLFSTFAVTECL 112  
 GPCR: 14 ALQTTTNYLIVSLAVADLLVATLVMPVWVYLEVVGWKFSSRIHCDIFVTLDDMMCTASIL 73  
 NOV11: 113 LLVMSYDLYVAICHPLRYLAIMTW-RVCITLAVTSWTTGVLLSLIHLVLLPLPFCRPQ 171  
 GPCR: 74 NLCAISIDRYTAVAMPMLYNTRYSSKRRVTVMIAIVWVLSFTISCPMLFGLNNTDQNE-- 131  
 40 NOV11: 172 KIIYHFFCEILAVLKLACADTHINENMVLAGAISGLVGPLSTIVVSYMCILCAILQIQSRE 231  
 GPCR: 132 -CIIANPAF-----VVSIVSFYVPFIVTLLVYIKIYIVLRRRRKRV 173  
 NOV11: 232 VQRK 235 (SEQ ID NO.: 81)  
 45 GPCR: 174 NTKR 177 (SEQ ID NO.: 82)

The OR family of the GPCR superfamily is a group of related proteins specifically located at the ciliated surface of olfactory sensory neurons in the nasal epithelium that are involved in the initial steps of the olfactory signal transduction cascade. Accordingly, the NOV11 nucleic acid, polypeptide, antibodies and other compositions of the present invention can be used to detect nasal epithelial neuronal tissue.

Based on its relatedness to the known members of the OR family of the GPCR superfamily, NOV11 satisfies a need in the art by providing new diagnostic or therapeutic

compositions useful in the treatment of disorders associated with alterations in the expression of members of OR family-like proteins. Nucleic acids, polypeptides, antibodies, and other compositions of the present invention are useful in the treatment and/or diagnosis of a variety of diseases and pathologies, including by way of nonlimiting example, those involving neurogenesis, cancer and wound healing.

## NOV12

A NOV12 sequence according to the invention is a nucleic acid sequence encoding a polypeptide related to the human odorant receptor (OR) family of the G-protein coupled receptor (GPCR) superfamily of proteins. . A NOV12 nucleic acid was discovered by exon linking analysis of NOV2 (SEQ ID NO.: 3). A NOV12 nucleic acid and its encoded polypeptide includes the sequences shown in Table 49. The disclosed nucleic acid (SEQ ID NO:23) is 1,014 nucleotides in length and contains an open reading frame (ORF) that begins with an ATG initiation codon at nucleotides 55-57 and ends with a TGA stop codon at nucleotides 985-987. The representative ORF encodes a 310 amino acid polypeptide (SEQ ID NO:24). Putative untranslated regions up- and downstream of the coding sequence are underlined in SEQ ID NO: 23.

**TABLE 49.**

TAAACACTTCTCCTAAACCATGAGCATTAACTTGATTTCCTCTGTCATAGGGATATG  
GGGGACAATATAACATCCATCACAGAGTTCCTCCTACTGGGATTTCCTCGTTGGCCCA  
AGGATTCAGATGCTCCTCTTTGGGCTCTTCTCCCTGTTCTACGTCTTCACCCTGCTGG  
GGAACGGGACCATACTGGGGCTCATCTCACTGGACTCCAGACTGCACGCCCCCATGT  
ACTTCTTCCTCTCACACCTGGCGGTGCGTACATCGCCTACGCCTGCAACACGGTGC  
25 CCCGGATGCTGGTGAACCTCCTGCATCCAGCCAAGCCCATCTCCTTTGCGGGCCGCA  
TGATGCAGACCTTTCTGTTTTCCACTTTTGCTGTACAGAATGTCTCCTCCTGGTGGT  
GATGTCCTATGATCTGTACGTGGCCATCTGCCACCCCCTCCGATATTTGGCCATCATG  
ACCTGGAGAGTCTGCATCACCTCGCGGTGACTTCCTGGACCACTGGAGTCCTTTTA  
TCCTTGATTCATCTTGTGTTACTTCTACCTTTACCCTTCTGTAGGCCCCAGAAAATTT  
30 ATCACTTTTGTGAAATCTTGGCTGTTCTCAAACCTTGCCTGTGCAGATACCCACAT  
CAATGAGAACATGGTCTTGGCCGGAGCAATTTCTGGGCTGGTGGGACCCTTGTCCAC  
AATTGTAGTTTCATATATGTGCATCCTCTGTGCTATCCTTCAGATCCAATCAAGGGAA  
GTTGAGAGGAAAGCCTTCTGCACCTGCTTCTCCACCTCTGTGTGATTGGACTCTTTT  
ATGGCACAGCCATTATCATGTATGTTGGACCCAGATATGGGAACCCCAAGGAGCAG  
35 AAGAAATATCTCCTGCTGTTTCACAGCCTCTTTAATCCCATGCTCAATCCCCTTATCT  
GTAGTCTTAGGAACTCAGAAGTGAAGAATACTTTGAAGAGAGTGCTGGGAGTAGAA  
AGGGCTTTATGAAAAGGATTATGGCATTGTGACTGACAA (SEQ ID NO.: 23)

MGDNITSITEFLLLGFVPVGPRIQMLLFGLFSLFYVFTLLGNGTILGLISLDSRLHAPMYFFLS  
HLAVVDIAIYACNTVPRMLVNLLHPAKPISFAGRMMQTFLFSTFAVTECLLLVMSYDLY  
VAICHPLRYLAIMTWRCITLAVTSWTTGVLLSLIHLVLLLPLPFCRPQKIYHFFCEILAVL  
KLACADTHINENMVLAGAISGLVGPLSTIVVSYMCILCAILQIQSREVQRKAFCTCFSHLC  
5 VIGLFYGTAIMYVGPRYGNPKEQKKYLLLFHSLFNPMLNPLICSLRNSEVKNTLKRVLG  
VERAL (SEQ ID NO.: 24)

The OR family of the GPCR superfamily is a group of related proteins specifically located at the ciliated surface of olfactory sensory neurons in the nasal epithelium and are involved in the initial steps of the olfactory signal transduction cascade. Accordingly, the NOV12 nucleic acid, polypeptide, antibodies and other compositions of the present invention can be used to detect nasal epithelial neuronal tissue.

The NOV12 polypeptide has a high degree of homology (approximately 99% identity) to a human olfactory receptor (OLFR) (EMBL Accession No.:095047), as is shown in Table 50.

15 The NOV12 polypeptide also has a high degree of homology (approximately 99% identity) to NOV2, as is shown in Table 51. Overall amino acid sequence identity within the mammalian OR family ranges from 45% to >80%. OR genes that are 80% or more identical to each other at the amino acid level are considered by convention to belong to the same subfamily. See *Dryer and Berghard, Trends in Pharmacological Sciences*, 1999, 20:413. Therefore, NOV12 and NOV2 are

20 two members of the same OR subfamily. OR proteins have seven transmembrane  $\alpha$ -helices separated by three extracellular and three cytoplasmic loops, with an extracellular amino-terminus and a cytoplasmic carboxy-terminus. Multiple sequence alignment suggests that the ligand-binding domain of the ORs is between the second and sixth transmembrane domains. NOV12 is predicted to have a seven transmembrane region, and is similar in that region to a

25 representative GPCR, *e.g.* dopamine (GPCR) (GenBank Accession No.: P20288) as is shown in Table 52.

**TABLE 50.**

30	NOV12:	1	MGDNITSITEFLLLGFVPVGPRIQMLLFGLFSLFYVFTLLGNGTILGLISLDSRLHAPMYF	60
			*****	
	OLFR:	1	MGDNITSIREFLLLGFVPVGPRIQMLLFGLFSLFYVFTLLGNGTILGLISLDSRLHAPMYF	60
35	NOV12:	61	FLSHLAVVDIAIYACNTVPRMLVNLLHPAKPISFAGRMMQTFLFSTFAVTECLLLVMSYD	120
			*****	
	OLFR:	61	FLSHLAVVDIAIYACNTVPRMLVNLLHPAKPISFAGRMMQTFLFSTFAVTECLLLVMSYD	120
40	NOV12:	121	LYVAICHPLRYLAIMTWRCITLAVTSWTTGVLLSLIHLVLLLPLPFCRPQKIYHFFCEI	180
			*****	
	OLFR:	121	LYVAICHPLRYLAIMTWRCITLAVTSWTTGVLLSLIHLVLLLPLPFCRPQKIYHFFCEI	180
	NOV12:	181	LAVLKLACADTHINENMVLAGAISGLVGPLSTIVVSYMCILCAILQIQSREVQRKAFCTC	240

\*\*\*\*\*  
 5 OLFR: 181 LAVLKLACADTHINENMVLAGAISGLVGPLSTIVVSYMCILCAILQIQSREVQRKAFRTC 240  
 NOV12: 241 FSHLCVIGLFYGTAIMYVGPRYGNPKEQKKYLLLFHSLFNPMLNPLICSLRNSEVKNTL 300  
 \*\*\*\*\*  
 OLFR: 241 FSHLCVIGLVYGTAIMYVGPRYGNPKEQKKYLLLFHSLFNPMLNPLICSLRNSEVKNTL 300  
 NOV12: 301 KRVLGVERAL 310 (SEQ ID NO.: 24)  
 \*\*\*\*\*  
 10 OLFR: 301 KRVLGVERAL 310 (SEQ ID NO.: 89)  
 Where \* indicates identity.

**TABLE 51.**

15 NOV12: 1 MGDNITSITEFLLLGFVPVGPRIQMLLFGLFSLFYVFTLLGNGTILGLISLDSRLHAPMYF 60  
 \*\*\*\*\*  
 NOV2: 1 MGDNITSIREFLLLGFVPVGPRIQMLLFGLFSLFYVFTLLGNGTILGLISLDSRLHAPMYF 60  
 NOV12: 61 FLSHLAVVDIAYACNTVPRMLVNLLHPAKPISFAGRMMQTFLFSTFAVTECLLLVMSYD 120  
 \*\*\*\*\*  
 20 NOV2: 61 FLSHLAVVDIAYACNTVPRMLVNLLHPAKPISFAGRMMQTFLFSTFAVTECLLLVMSYD 120  
 NOV12: 121 LYVAICHPLRYLAIMTWRCITLAVTSWTTGVLLSLIHLVLLLPLPFCRPQKIYHFFCEI 180  
 \*\*\*\*\*  
 25 NOV2: 121 LYVAICHPLRYLAIMTWRCITLAVTSWTTGVLLSLIHLVLLLPLPFCRPQKIYHFFCEI 180  
 NOV12: 181 LAVLKLACADTHINENMVLAGAISGLVGPLSTIVVSYMCILCAILQIQSREVQRKAFCTC 240  
 \*\*\*\*\*  
 NOV2: 181 LAVLKLACADTHINENMVLAGAISGLVGPLSTIVVSYMCILCAILQIQSREVQRKAFRTC 240  
 30 NOV12: 241 FSHLCVIGLFYGTAIMYVGPRYGNPKEQKKYLLLFHSLFNPMLNPLICSLRNSEVKNTL 300  
 \*\*\*\*\*  
 NOV2: 241 FSHLCVIGLVYGTAIMYVGPRYGNPKEQKKYLLLFHSLFNPMLNPLICSLRNSEVKNTL 300  
 NOV12: 301 KRVLGVERAL 310 (SEQ ID NO.: 24)  
 \*\*\*\*\*  
 35 NOV2: 301 KRVLGVERAL 310 (SEQ ID NO.: 4)  
 Where \* indicates identity.

**TABLE 52.**

40 NOV12: 53 RLHAPMYFFLSHLAVVDIAYACNTVPRMLVNLLHPAKPISFAGRMMQTFLFSTFAVTECL 112  
 GPCR: 14 ALQTTTNYLIVSLAVADLLVATLVMPWVYVLEVVGWKFRIHCDIFVTLDVMMCTASIL 73  
 NOV12: 113 LLVMSYDLYVAICHPLRYLAIMTW-RVCITLAVTSWTTGVLLSLIHLVLLLPLPFCRPQ 171  
 45 GPCR: 74 NLCAISIDRYTAVAMPMLYNTRYSSKRRVTVMIAIVWVLSFTISCPMLFGLNNTDQNE-- 131  
 NOV12: 172 KIYHFFCEILAVLKLACADTHINENMVLAGAISGLVGPLSTIVVSYMCILCAILQIQSRE 231  
 GPCR: 132 -CIIANPAF-----VVISSIVSFYVPIVTLVYIKIYIVLRRRRKRV 173  
 NOV12: 232 VQRK 235 (SEQ ID NO.: 90)  
 50 GPCR: 174 NTKR 177 (SEQ ID NO.: 91)

The OR family of the GPCR superfamily is a group of related proteins specifically located at the ciliated surface of olfactory sensory neurons in the nasal epithelium that are involved in the initial steps of the olfactory signal transduction cascade. Accordingly, the NOV12



nucleic acid, polypeptide, antibodies and other compositions of the present invention can be used to detect nasal epithelial neuronal tissue.

Based on its relatedness to the known members of the OR family of the GPCR superfamily, NOV12 satisfies a need in the art by providing new diagnostic or therapeutic compositions useful in the treatment of disorders associated with alterations in the expression of members of OR family-like proteins. Nucleic acids, polypeptides, antibodies, and other compositions of the present invention are useful in the treatment and/or diagnosis of a variety of diseases and pathologies, including by way of nonlimiting example, those involving neurogenesis, cancer and wound healing.

### NOV13

A NOV13 sequence according to the invention is a nucleic acid sequence encoding a polypeptide related to the human odorant receptor (OR) family of the G-protein coupled receptor (GPCR) superfamily of proteins. A NOV13 nucleic acid and its encoded polypeptide includes the sequences shown in Table 53. The disclosed nucleic acid (SEQ ID NO:25) is 908 nucleotides in length and contains an open reading frame (ORF) that begins with an ATG initiation codon at nucleotides 75-77 and ends with a TAA stop codon at nucleotides 901-903. The representative ORF encodes a 270 amino acid polypeptide (SEQ ID NO:26). Putative untranslated regions up- and downstream of the coding sequence are underlined in SEQ ID NO: 25.

**TABLE 53.**

TGTATCTGGTCACGGTGCTGAGGAACCTGCTCAGCATCCTGGCTGTCAGCTCTGACT  
CCCACCCCCACACACCCATGTACTTCTTCCTCTCCAACCTGTGCTGGGCTGACATCG  
GTTTCACCTTGGCCACGGTTCCCAAGATGATTGTGGACATGGGGTCGCATAGCAGAG  
TCATCTCTTATGAGGGCTGCCTGACACAGATGTCTTTCTTTGTCCTTTTTGCATGTAT  
AGAAGACATGCTCCTGACTGTGATGGCCTATGACCAATTTGTGGCCATCTGTACCCC  
CCTGCACTACCCAGTCATCATGAATCCTCACCTCTGTGTCTTCTTAGTTTTGGTTTTCTT  
TTTTCTTAGCCTGTTGGATTCCCAGCTGCACAGTTGGATTGTGTTACAATTCACCTT  
CTTCAAGAATGTGGAAATCTCTAATTTTTTCTGTGATCCATCTCAACTTCTCAACCTT  
GCCTGTTCTGACGGCATCATCAATAGCATATTCATATATTTAGATAGTATTCTGTTCA  
GTTTTCTTCCCATTTTCAGGGATCCTTTTGTCTTACTATAAAATTGTCCCCTCCATTCTA  
AGAATTTTCATCGTCAGATGGGAAGTATAAAGCCTTCTCCATCTGTGGCTCTCACCTG  
GCAGTTGTTTGCTTATTTTATGGAACAGGCATTGGCGTGTACCTAACTTCAGCTGTGT  
CACCACCCCCCAGGAATGGTGTGGTGGCGTCAGTGATGTATGCTGTGGTCACCCCCA  
TGCTGAACCCTTTTCATCTACAGCCTGAGAAACAGGGATATACAAAGTGTCCTGCGGA  
GGCTGTGCAGCAGAACAGTCGAATCTCATGATATGTTCCATCCTTTTTCTTGTGTGGG  
TGAGAAAGGGCAACCACATTAAATCTCTACATCTGTAAATCCT (SEQ ID NO.: 25)



5 MYFFLSNLCWADIGFTLATVPKMIVDMGSHSRVISYEGCLTQMSFFVLFACIEDMLLTV  
MAYDQFVAICHPLHYPVIMNPHLCVFLVLVSFFLSLLDSQLHSWIVLQFTFFKNVEISNFF  
CDPSQLNLACSDGIINSIFIYLDLSILFSFLPISGILLSYYKIVPSILRISSSDGKYKAFCGSHL  
AVVCLFYGTGIGVYLTSAVSPPPRNGVVASVMYAVVTPMLNPFYSLRNRDIQSVLRRRLC  
SRTVESHDMFHPFSCVGEKGQPH (SEQ ID NO.: 26)

The OR family of the GPCR superfamily is a group of related proteins specifically located at the ciliated surface of olfactory sensory neurons in the nasal epithelium and are involved in the initial steps of the olfactory signal transduction cascade. Accordingly, the NOV13 nucleic acid, polypeptide, antibodies and other compositions of the present invention can be used to detect nasal epithelial neuronal tissue.

The NOV13 polypeptide has homology (approximately 73% identity, 83% similarity) to a human olfactory receptor (OLFR) (EMBL Accession No.: Q9UPJ1), as is shown in Table 54. Overall amino acid sequence identity within the mammalian OR family ranges from 45% to >80%. OR genes that are 80% or more identical to each other at the amino acid level are considered by convention to belong to the same subfamily. See *Dryer and Berghard, Trends in Pharmacological Sciences*, 1999, 20:413. OR proteins have seven transmembrane  $\alpha$ -helices separated by three extracellular and three cytoplasmic loops, with an extracellular amino-terminus and a cytoplasmic carboxy-terminus. Multiple sequence alignment suggests that the ligand-binding domain of the ORs is between the second and sixth transmembrane domains. NOV13 is predicted to have a seven transmembrane region, and is similar in that region to a representative GPCR, e.g. dopamine (GPCR) (GenBank Accession No.: P20288) as is shown in Table 55.

**TABLE 54**

NOV12:	1	MYFFLSNLCWADIGFTLATVPKMIVDMGSHSRVISYEGCLTQMSFFVLFACIEDMLLTV	60
		*****	
OLFR:	1	MYFFLSNLSLADIGFTSTTVPKMIVDMQTHSRVISYEGCLTQMSFFVLFACMDDMLLSVM	60
NOV12:	61	AYDQFVAICHPLHYPVIMNPHLCVFLVLVSFFLSLLDSQLHSWIVLQFTFFKNVEISNFF	120
		***+*****	
OLFR:	61	AYDRFVAICHPLHYRIIMNPRLCGFLILLSFFISLLDSQLHNLIMLQLTCEKDVDISNFF	120
NOV12:	121	CDPSQLNLACSDGIINSIFIYLDLSILFSFLPISGILLSYYKIVPSILRISSSDGKYKAF	180
		*****+* ** *	
OLFR:	121	CDPSQLLHLRCSDTFINEMVIYFMGAIFGCLPISGILFSYYKIVSPILRVPTSDGKYKAF	180
NOV12:	181	SICGSHLAVVCLFYGTGIGVYLTSAVSPPPRNGVVASVMYAVVTPMLNPFYSLRNRDIQ	240
		* *****+ **++** *	
OLFR:	181	STCGSHLAVVCLFYGTGLVGYLSSAVLPSPRKSMVASVMYTVVTPMLNPFYSLRNKDIQ	240
NOV12:	241	SVLRRRLCSRTVESHDMFHPFSCVG	263 (SEQ ID NO.: 24)

\* \* \* \* \* + + \* \* \* + \*

OLFR: 241 SALCRLHGRIIKSHHL-HPFCYMG 263 (SEQ ID NO.: 92)  
Where \* indicates identity and + indicates similarity.

5                    **TABLE 55**

NOV13: 1 MYFFLSNLCWADIGFTLATVPKMIIVDMGSHSRVISYEGCLTQMSFFVLFIACIEDMLLTVM 60  
GPCR: 19 TNYLIVSLAVADLLVATLVMPWVYVLEVVGEWKFSRIHCDIFVTLDVMMCTASILNLCAI 78

10 NOV13:61 AYDQFVAICHPLHPVIMNPHLCVFLVLVSFFLSLLDSQLHSHWIVLQFTF-FKNVEISNF 119  
GPCR: 79 SIDRYTAVAMPMLYNTRYSSKRRVTVMIAIVWVLSF-----TISCPMLFGLNNTDQNECI 133

NOV13:120 FCDPSQLNLACSDGIINSIFIYLDLSILFSFLPISGILLSYYKIVPSILRISS 173  
(SEQ ID NO.: 93)  
GPCR: 134 IANPAFVV-----YSSIVSFYVPPFIVTLLVYIKIYIVLRRRRKR 172  
15 (SEQ ID NO.: 94)

The OR family of the GPCR superfamily is a group of related proteins specifically located at the ciliated surface of olfactory sensory neurons in the nasal epithelium that are involved in the initial steps of the olfactory signal transduction cascade. Accordingly, the NOV13 nucleic acid, polypeptide, antibodies and other compositions of the present invention can be used to detect nasal epithelial neuronal tissue.

Based on its relatedness to the known members of the OR family of the GPCR superfamily, NOV13 satisfies a need in the art by providing new diagnostic or therapeutic compositions useful in the treatment of disorders associated with alterations in the expression of members of OR family-like proteins. Nucleic acids, polypeptides, antibodies, and other compositions of the present invention are useful in the treatment and/or diagnosis of a variety of diseases and pathologies, including by way of nonlimiting example, those involving neurogenesis, cancer and wound healing.

30                    Table 56 shows a multiple sequence alignment of NOV1-13 polypeptides with the known human olfactory receptor 10J1 (GenBank Accession No.: P30954), indicating the homology between the present invention and known members of a protein family.

**TABLE 56.**

35 NOV4        -----MRGFNKT--TVVTQFILVGFSSLGELQ--LLLFVIFLLLYLTILVANVTIMA  
NOV3        -----MRGFNKT--TVVTQFILVGFSSLGELQ--LLLFVIFLLLYLTILVANVTIMA  
OR\_10J1 MLLCFRFGNQSMKRENFTLITDFVFQGFSSFHEQQ--ITLFGVFLALYILTLAGNIIIVT  
NOV10       -----MNPANHSQVAGFVLLGLSQVWELR--FVFFTVFSVAVYFMTVVGNNLIVV  
40 NOV12       -----MGDNITSIT-EFLLLGFPVGPRIQ--MLLFGLFSLFYVFTLLGNGTILG  
NOV11       -----MGDNITSIT-EFLLLGFPVGPRIQ--MLLFGLFSLFYVFTLLGNGTILG  
NOV2        -----MGDNITSIR-EFLLLGFPVGPRIQ--MLLFGLFSLFYVFTLLGNGTILG  
NOV9        -----MGDVNQSVASDFILVGLFSGSGSR--QLLFSLVAVMFVIGLLGNTVLLF  
NOV8        -----MGDVNQSVASDFILVGLFSGSGSR--QLLFSLVAVMFVIGLLGNTVLLF  
NOV1\_       -----MEGKNQTNISEFLLLGFSWQQQQ--VLLFALFLCLYLTGLFGNLLILL

[illegible]

\* : \* . \* : : . : \* . : : . \* \*\*

	NOV4	LVYSLRNKEVKTALKR-----VLGMPVATKMS-----	(SEQ ID NO.: 8)
	NOV3	LVYSLRNKEVKTALKR-----VLGMPVATKMS-----	(SEQ ID NO.: 6)
5	R_10J1	VVYTLRNKEVKDALCR-----AVGG----KFS-----	(SEQ ID NO.: 95)
	NOV10	AIYTLRNKEVIMAMKKLWRRKKDPIGPLEHRPLH-----	(SEQ ID NO.: 20)
	NOV12	LICSLRNSEVKNTLKR-----VLG--VERAL-----	(SEQ ID NO.: 24)
	NOV11	LICSLRNSEVKNTLKR-----VLG--VERAL-----	(SEQ ID NO.: 22)
	NOV2	LICSLRNSEVKNTLKR-----VLG--VERAL-----	(SEQ ID NO.: 4)
10	NOV9	LIYSLRNPEVWMALVK-----VLSRAGLRQMCMTT-----	(SEQ ID NO.: 18)
	NOV8	LIYSLRNPEVWMALVK-----VLSRAGLRQMC-----	(SEQ ID NO.: 16)
	NOV1	FIYSLRNNELKGTLLKKTLSRPGAVAHACNPSTLGGRGGWIMRSGDRDHPG	(SEQ ID NO.: 2)
	NOV6	FILSLGKRDIQSVLRRRLCSRTVESHDMFHPFSCVGEKGQPH-----	(SEQ ID NO.: 12)
	NOV5	FIYSLGKRDIQSVLRRRLCSRTVESHDMFHPFSCVG-----	(SEQ ID NO.: 10)
15	NOV13	FIYSLRNRDIQSVLRRRLCSRTVESHDMFHPFSCVGEKGQPH-----	(SEQ ID NO.: 26)
	NOV7	FIYSLRNRDIQSALRRLRSR-----	(SEQ ID NO.: 14)

: : \* : : : . : :

20 Where “\*” indicates a single, fully conserved residue, “:” indicates conservation of strong groups, and “.” indicates conservation of weak groups, and OR\_10J1 is the known human olfactory receptor 10J1 (GenBank Accession No.: P30954).

25 The nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in disorders of the neuro-olfactory system, such as those induced by trauma, surgery and/or neoplastic disorders. For example, a cDNA encoding the olfactory receptor protein may be useful in gene therapy for treating such disorders, and the olfactory receptor protein may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from disorders of the neuro-olfactory system. The novel nucleic acids

30 encoding olfactory receptor protein, and the olfactory receptor protein of the invention, or fragments thereof, may further be useful in the treatment of adenocarcinoma; lymphoma; prostate cancer; uterus cancer, immune response, AIDS, asthma, Crohn's disease, multiple sclerosis, treatment of Albright hereditary osteodystrophy, development of powerful assay system for functional analysis of various human disorders which will help in understanding of pathology of

35 the disease, and development of new drug targets for various disorders. They may also be used in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. These materials are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods.

40

#### NOVX Nucleic Acids

The nucleic acids of the invention include those that encode a NOVX polypeptide or protein. As used herein, the terms polypeptide and protein are interchangeable.

09999999-070404  
T05040-999999

In some embodiments, a NOVX nucleic acid encodes a mature NOVX polypeptide. As used herein, a “mature” form of a polypeptide or protein described herein relates to the product of a naturally occurring polypeptide or precursor form or proprotein. The naturally occurring polypeptide, precursor or proprotein includes, by way of nonlimiting example, the full-length gene product, encoded by the corresponding gene. Alternatively, it may be defined as the polypeptide, precursor or proprotein encoded by an open reading frame described herein. The product “mature” form arises, again by way of nonlimiting example, as a result of one or more naturally occurring processing steps that may take place within the cell in which the gene product arises. Examples of such processing steps leading to a “mature” form of a polypeptide or protein include the cleavage of the N-terminal methionine residue encoded by the initiation codon of an open reading frame, or the proteolytic cleavage of a signal peptide or leader sequence. Thus a mature form arising from a precursor polypeptide or protein that has residues 1 to N, where residue 1 is the N-terminal methionine, would have residues 2 through N remaining after removal of the N-terminal methionine. Alternatively, a mature form arising from a precursor polypeptide or protein having residues 1 to N, in which an N-terminal signal sequence from residue 1 to residue M is cleaved, would have the residues from residue M+1 to residue N remaining. Further as used herein, a “mature” form of a polypeptide or protein may arise from a step of post-translational modification other than a proteolytic cleavage event. Such additional processes include, by way of non-limiting example, glycosylation, myristoylation or phosphorylation. In general, a mature polypeptide or protein may result from the operation of only one of these processes, or a combination of any of them.

Among the NOVX nucleic acids is the nucleic acid whose sequence is provided in SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23 or 25, or a fragment thereof. Additionally, the invention includes mutant or variant nucleic acids of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23 or 25, or a fragment thereof, any of whose bases may be changed from the corresponding bases shown in SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23 or 25, while still encoding a protein that maintains at least one of its NOVX-like activities and physiological functions (*i.e.*, modulating angiogenesis, neuronal development). The invention further includes the complement of the nucleic acid sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23 or 25, including fragments, derivatives, analogs and homologs thereof. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications.

10

15

20

30



complement of any of this nucleotide sequence, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23 or 25, as a hybridization probe, NOVX nucleic acid sequences can be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook *et al.*, eds., MOLECULAR CLONING: A LABORATORY MANUAL 2<sup>nd</sup> Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989; and Ausubel, *et al.*, eds., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993.)

A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to NOVX nucleotide sequences can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

As used herein, the term "oligonucleotide" refers to a series of linked nucleotide residues, which oligonucleotide has a sufficient number of nucleotide bases to be used in a PCR reaction. A short oligonucleotide sequence may be based on, or designed from, a genomic or cDNA sequence and is used to amplify, confirm, or reveal the presence of an identical, similar or complementary DNA or RNA in a particular cell or tissue. Oligonucleotides comprise portions of a nucleic acid sequence having about 10 nt, 50 nt, or 100 nt in length, preferably about 15 nt to 30 nt in length. In one embodiment, an oligonucleotide comprising a nucleic acid molecule less than 100 nt in length would further comprise at least 6 contiguous nucleotides of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23 or 25, or a complement thereof. Oligonucleotides may be chemically synthesized and may be used as probes.

In another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule that is a complement of the nucleotide sequence shown in SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23 or 25, or a portion of this nucleotide sequence. A nucleic acid molecule that is complementary to the nucleotide sequence shown in SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23 or 25 is one that is sufficiently complementary to the nucleotide sequence shown in SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23 or 25 that it can hydrogen bond with little or no mismatches to the nucleotide sequence shown in SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23 or 25, thereby forming a stable duplex.



As used herein, the term “complementary” refers to Watson-Crick or Hoogsteen base pairing between nucleotide units of a nucleic acid molecule, and the term “binding” means the physical or chemical interaction between two polypeptides or compounds or associated polypeptides or compounds or combinations thereof. Binding includes ionic, non-ionic, Von der  
5 Waals, hydrophobic interactions, etc. A physical interaction can be either direct or indirect. Indirect interactions may be through or due to the effects of another polypeptide or compound. Direct binding refers to interactions that do not take place through, or due to, the effect of another polypeptide or compound, but instead are without other substantial chemical intermediates.

Moreover, the nucleic acid molecule of the invention can comprise only a portion of the  
10 nucleic acid sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23 or 25, *e.g.*, a fragment that can be used as a probe or primer, or a fragment encoding a biologically active portion of NOVX. Fragments provided herein are defined as sequences of at least 6 (contiguous) nucleic acids or at least 4 (contiguous) amino acids, a length sufficient to allow for specific hybridization in the case of nucleic acids or for specific recognition of an epitope in the case of  
15 amino acids, respectively, and are at most some portion less than a full length sequence. Fragments may be derived from any contiguous portion of a nucleic acid or amino acid sequence of choice. Derivatives are nucleic acid sequences or amino acid sequences formed from the native compounds either directly or by modification or partial substitution. Analogs are nucleic acid sequences or amino acid sequences that have a structure similar to, but not identical to, the  
20 native compound but differs from it in respect to certain components or side chains. Analogs may be synthetic or from a different evolutionary origin and may have a similar or opposite metabolic activity compared to wild type.

Derivatives and analogs may be full length or other than full length, if the derivative or analog contains a modified nucleic acid or amino acid, as described below. Derivatives or  
25 analogs of the nucleic acids or proteins of the invention include, but are not limited to, molecules comprising regions that are substantially homologous to the nucleic acids or proteins of the invention, in various embodiments, by at least about 70%, 80%, 85%, 90%, 95%, 98%, or even 99% identity (with a preferred identity of 80-99%) over a nucleic acid or amino acid sequence of identical size or when compared to an aligned sequence in which the alignment is done by a  
30 computer homology program known in the art, or whose encoding nucleic acid is capable of hybridizing to the complement of a sequence encoding the aforementioned proteins under stringent, moderately stringent, or low stringent conditions. See *e.g.* Ausubel, *et al.*, CURRENT

A “homologous nucleic acid sequence” or “homologous amino acid sequence,” or variations thereof, refer to sequences characterized by a homology at the nucleotide level or amino acid level as discussed above. Homologous nucleotide sequences encode those sequences coding for isoforms of a NOVX polypeptide. Isoforms can be expressed in different tissues of the same organism as a result of, for example, alternative splicing of RNA. Alternatively, isoforms can be encoded by different genes. In the present invention, homologous nucleotide sequences include nucleotide sequences encoding for a NOVX polypeptide of species other than humans, including, but not limited to, mammals, and thus can include, *e.g.*, mouse, rat, rabbit, dog, cat, cow, horse, and other organisms. Homologous nucleotide sequences also include, but are not limited to, naturally occurring allelic variations and mutations of the nucleotide sequences set forth herein. A homologous nucleotide sequence does not, however, include the nucleotide sequence encoding human NOVX protein. Homologous nucleic acid sequences include those nucleic acid sequences that encode conservative amino acid substitutions (see below) in SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24 or 26, as well as a polypeptide having NOVX activity. Biological activities of the NOVX proteins are described below. A homologous amino acid sequence does not encode the amino acid sequence of a human NOVX polypeptide.

The nucleotide sequence determined from the cloning of the human NOVX gene allows for the generation of probes and primers designed for use in identifying and/or cloning NOVX homologues in other cell types, *e.g.*, from other tissues, as well as NOVX homologues from other mammals. The probe/primer typically comprises a substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 25, 50, 100, 150, 200, 250, 300, 350 or 400 or more consecutive sense strand nucleotide sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23 or 25; or an anti-sense strand nucleotide sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23 or 25; or of a naturally occurring mutant of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23 or 25.

Probes based on the human NOVX nucleotide sequence can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In various embodiments, the probe further comprises a label group attached thereto, *e.g.*, the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissue which misexpress a NOVX protein, such as by measuring a level of a NOVX-encoding nucleic acid in a sample of cells from a subject *e.g.*, detecting NOVX mRNA levels or determining whether a genomic NOVX gene has been mutated or deleted.

A "polypeptide having a biologically active portion of NOVX" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the present invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. A nucleic acid fragment encoding a "biologically active portion of NOVX" can be prepared by isolating a portion of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23 or 25 that encodes a polypeptide having a NOVX biological activity (biological activities of the NOVX proteins are described below), expressing the encoded portion of NOVX protein (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of NOVX. For example, a nucleic acid fragment encoding a biologically active portion of NOVX can optionally include an ATP-binding domain. In another embodiment, a nucleic acid fragment encoding a biologically active portion of NOVX includes one or more regions.

#### NOVX Variants

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequences shown in SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23 or 25 due to the degeneracy of the genetic code. These nucleic acids thus encode the same NOVX protein as that encoded by the nucleotide sequence shown in SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23 or 25 *e.g.*, the polypeptide of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24 or 26. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24 or 26.

In addition to the human NOVX nucleotide sequence shown in SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23 or 25, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of NOVX may exist

within a population (*e.g.*, the human population). Such genetic polymorphism in the NOVX gene may exist among individuals within a population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame encoding a NOVX protein, preferably a mammalian NOVX protein. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of the NOVX gene. Any and all such nucleotide variations and resulting amino acid polymorphisms in NOVX that are the result of natural allelic variation and that do not alter the functional activity of NOVX are intended to be within the scope of the invention.

Moreover, nucleic acid molecules encoding NOVX proteins from other species, and thus that have a nucleotide sequence that differs from the human sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23 or 25 are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and homologues of the NOVX cDNAs of the invention can be isolated based on their homology to the human NOVX nucleic acids disclosed herein using the human cDNAs, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions. For example, a soluble human NOVX cDNA can be isolated based on its homology to human membrane-bound NOVX. Likewise, a membrane-bound human NOVX cDNA can be isolated based on its homology to soluble human NOVX.

Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 6 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23 or 25. In another embodiment, the nucleic acid is at least 10, 25, 50, 100, 250, 500 or 750 nucleotides in length. In another embodiment, an isolated nucleic acid molecule of the invention hybridizes to the coding region. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% homologous to each other typically remain hybridized to each other.

Homologs (*i.e.*, nucleic acids encoding NOVX proteins derived from species other than human) or other related sequences (*e.g.*, paralogs) can be obtained by low, moderate or high stringency hybridization with all or a portion of the particular human sequence as a probe using methods well known in the art for nucleic acid hybridization and cloning.

As used herein, the phrase "stringent hybridization conditions" refers to conditions under which a probe, primer or oligonucleotide will hybridize to its target sequence, but to no other

sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures than shorter sequences. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength and pH. The  $T_m$  is the temperature (under defined ionic strength, pH and nucleic acid concentration) at which 50% of the probes complementary to the target sequence hybridize to the target sequence at equilibrium. Since the target sequences are generally present at excess, at  $T_m$ , 50% of the probes are occupied at equilibrium. Typically, stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes, primers or oligonucleotides (e.g., 10 nt to 50 nt) and at least about 60°C for longer probes, primers and oligonucleotides. Stringent conditions may also be achieved with the addition of destabilizing agents, such as formamide.

Stringent conditions are known to those skilled in the art and can be found in CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. Preferably, the conditions are such that sequences at least about 65%, 70%, 75%, 85%, 90%, 95%, 98%, or 99% homologous to each other typically remain hybridized to each other. A non-limiting example of stringent hybridization conditions is hybridization in a high salt buffer comprising 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 mg/ml denatured salmon sperm DNA at 65°C. This hybridization is followed by one or more washes in 0.2X SSC, 0.01% BSA at 50°C. An isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23 or 25 corresponds to a naturally occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein).

In a second embodiment, a nucleic acid sequence that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23 or 25, or fragments, analogs or derivatives thereof, under conditions of moderate stringency is provided. A non-limiting example of moderate stringency hybridization conditions are hybridization in 6X SSC, 5X Denhardt's solution, 0.5% SDS and 100 mg/ml denatured salmon sperm DNA at 55°C, followed by one or more washes in 1X SSC, 0.1% SDS at 37°C. Other conditions of moderate stringency that may be used are well known in the art. See, e.g., Ausubel

*et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990, GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY.

In a third embodiment, a nucleic acid that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23 or 25, or fragments, analogs or derivatives thereof, under conditions of low stringency, is provided. A non-limiting example of low stringency hybridization conditions are hybridization in 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 mg/ml denatured salmon sperm DNA, 10% (wt/vol) dextran sulfate at 40°C, followed by one or more washes in 2X SSC, 25 mM Tris-HCl (pH 7.4), 5 mM EDTA, and 0.1% SDS at 50°C. Other conditions of low stringency that may be used are well known in the art (*e.g.*, as employed for cross-species hybridizations). See, *e.g.*, Ausubel *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990, GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY; Shilo and Weinberg, 1981, *Proc Natl Acad Sci USA* 78: 6789-6792.

#### Conservative mutations

In addition to naturally-occurring allelic variants of the NOVX sequence that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into the nucleotide sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23 or 25, thereby leading to changes in the amino acid sequence of the encoded NOVX protein, without altering the functional ability of the NOVX protein. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in the sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23 or 25. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of NOVX without altering the biological activity, whereas an "essential" amino acid residue is required for biological activity. For example, amino acid residues that are conserved among the NOVX proteins of the present invention, are predicted to be particularly unamenable to alteration.

Another aspect of the invention pertains to nucleic acid molecules encoding NOVX proteins that contain changes in amino acid residues that are not essential for activity. Such NOVX proteins differ in amino acid sequence from SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20,



22, 24 or 26, yet retain biological activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 75% homologous to the amino acid sequence of SEQ ID NO: 2, 4, 6, or 8. Preferably, the protein encoded by the nucleic acid is at least about 80% homologous to SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24 or 26, more preferably at least about 90%, 95%, 98%, and most preferably at least about 99% homologous to SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24 or 26.

An isolated nucleic acid molecule encoding a NOVX protein homologous to the protein of can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23 or 25, such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein.

Mutations can be introduced into the nucleotide sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23 or 25 by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in NOVX is replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a NOVX coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for NOVX biological activity to identify mutants that retain activity. Following mutagenesis of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23 or 25 the encoded protein can be expressed by any recombinant technology known in the art and the activity of the protein can be determined.

In one embodiment, a mutant NOVX protein can be assayed for (1) the ability to form protein:protein interactions with other NOVX proteins, other cell-surface proteins, or biologically active portions thereof, (2) complex formation between a mutant NOVX protein and a NOVX receptor; (3) the ability of a mutant NOVX protein to bind to an intracellular target protein or biologically active portion thereof; (*e.g.*, avidin proteins); (4) the ability to bind NOVX protein; or (5) the ability to specifically bind an anti-NOVX protein antibody.

#### Antisense NOVX Nucleic Acids

Another aspect of the invention pertains to isolated antisense nucleic acid molecules that are hybridizable to or complementary to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23 or 25, or fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein, *e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. In specific aspects, antisense nucleic acid molecules are provided that comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire NOVX coding strand, or to only a portion thereof. Nucleic acid molecules encoding fragments, homologs, derivatives and analogs of a NOVX protein of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24 or 26 or antisense nucleic acids complementary to a NOVX nucleic acid sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23 or 25 are additionally provided.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding NOVX. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues (*e.g.*, the protein coding region of human NOVX corresponds to SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24 or 26). In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding NOVX. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding NOVX disclosed herein (*e.g.*, SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23 or 25), antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick or Hoogsteen base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of NOVX mRNA, but

more preferably is an oligonucleotide that is antisense to only a portion of the coding or noncoding region of NOVX mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of NOVX mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis or enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used.

Examples of modified nucleotides that can be used to generate the antisense nucleic acid include: 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (*v*), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (*v*), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a NOVX protein to thereby inhibit expression of the protein, *e.g.*, by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid

1 molecule that binds to DNA duplexes, through specific interactions in the major groove of the  
double helix. An example of a route of administration of antisense nucleic acid molecules of the  
invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules  
can be modified to target selected cells and then administered systemically. For example, for  
5 systemic administration, antisense molecules can be modified such that they specifically bind to  
receptors or antigens expressed on a selected cell surface, *e.g.*, by linking the antisense nucleic  
acid molecules to peptides or antibodies that bind to cell surface receptors or antigens. The  
antisense nucleic acid molecules can also be delivered to cells using the vectors described herein.  
To achieve sufficient intracellular concentrations of antisense molecules, vector constructs in  
10 which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III  
promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an  
 $\alpha$ -anomeric nucleic acid molecule. An  $\alpha$ -anomeric nucleic acid molecule forms specific  
double-stranded hybrids with complementary RNA in which, contrary to the usual  $\beta$ -units, the  
15 strands run parallel to each other (Gaultier *et al.* (1987) *Nucleic Acids Res* 15: 6625-6641). The  
antisense nucleic acid molecule can also comprise a 2'-*o*-methylribonucleotide (Inoue *et al.*  
(1987) *Nucleic Acids Res* 15: 6131-6148) or a chimeric RNA -DNA analogue (Inoue *et al.* (1987)  
*FEBS Lett* 215: 327-330).

Such modifications include, by way of nonlimiting example, modified bases, and nucleic  
20 acids whose sugar phosphate backbones are modified or derivatized. These modifications are  
carried out at least in part to enhance the chemical stability of the modified nucleic acid, such  
that they may be used, for example, as antisense binding nucleic acids in therapeutic applications  
in a subject.

## 25 **NOVX Ribozymes and PNA moieties**

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme.  
Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a  
single-stranded nucleic acid, such as a mRNA, to which they have a complementary region.  
Thus, ribozymes (*e.g.*, hammerhead ribozymes (described in Haselhoff and Gerlach (1988)  
30 *Nature* 334:585-591)) can be used to catalytically cleave NOVX mRNA transcripts to thereby  
inhibit translation of NOVX mRNA. A ribozyme having specificity for a NOVX-encoding  
nucleic acid can be designed based upon the nucleotide sequence of a NOVX DNA disclosed

herein (*i.e.*, SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23 or 25). For example, a derivative of a Tetrahymena L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a NOVX-encoding mRNA. See, *e.g.*, Cech *et al.* U.S. Pat. No. 4,987,071; and Cech *et al.* U.S. Pat. No. 5,116,742. Alternatively, NOVX mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, *e.g.*, Bartel *et al.*, (1993) *Science* 261:1411-1418.

Alternatively, NOVX gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the NOVX (*e.g.*, the NOVX promoter and/or enhancers) to form triple helical structures that prevent transcription of the NOVX gene in target cells. See generally, Helene. (1991) *Anticancer Drug Des.* 6: 569-84; Helene. *et al.* (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher (1992) *Bioassays* 14: 807-15.

In various embodiments, the nucleic acids of NOVX can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (see Hyrup *et al.* (1996) *Bioorg Med Chem* 4: 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, *e.g.*, DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup *et al.* (1996) above; Perry-O'Keefe *et al.* (1996) *PNAS* 93: 14670-675.

PNAs of NOVX can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, *e.g.*, inducing transcription or translation arrest or inhibiting replication. PNAs of NOVX can also be used, *e.g.*, in the analysis of single base pair mutations in a gene by, *e.g.*, PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, *e.g.*, S1 nucleases (Hyrup B. (1996) above); or as probes or primers for DNA sequence and hybridization (Hyrup *et al.* (1996), above; Perry-O'Keefe (1996), above).

In another embodiment, PNAs of NOVX can be modified, *e.g.*, to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of

PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of NOVX can be generated that may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, e.g., RNase H and DNA polymerases, to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup (1996) above). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup (1996) above and Finn *et al.* (1996) *Nucl Acids Res* 24: 3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry, and modified nucleoside analogs, e.g., 5'-(4-methoxytrityl) amino-5'-deoxy-thymidine phosphoramidite, can be used between the PNA and the 5' end of DNA (Mag *et al.* (1989) *Nucl Acid Res* 17: 5973-88). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn *et al.* (1996) above). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment. See, Petersen *et al.* (1975) *Bioorg Med Chem Lett* 5: 1119-11124.

In other embodiments, the oligonucleotide may include other appended groups such as peptides (e.g., for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, e.g., Letsinger *et al.*, 1989, *Proc. Natl. Acad. Sci. U.S.A.* 86:6553-6556; Lemaitre *et al.*, 1987, *Proc. Natl. Acad. Sci.* 84:648-652; PCT Publication No. W088/09810) or the blood-brain barrier (see, e.g., PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization triggered cleavage agents (See, e.g., Krol *et al.*, 1988, *BioTechniques* 6:958-976) or intercalating agents. (See, e.g., Zon, 1988, *Pharm. Res.* 5: 539-549). To this end, the oligonucleotide may be conjugated to another molecule, e.g., a peptide, a hybridization triggered cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, etc.

### NOVX Polypeptides

A NOVX polypeptide of the invention includes the NOVX-like protein whose sequence is provided in SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24 or 26. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24 or 26 while



still encoding a protein that maintains its NOVX-like activities and physiological functions, or a functional fragment thereof. In some embodiments, up to 20% or more of the residues may be so changed in the mutant or variant protein. In some embodiments, the NOVX polypeptide according to the invention is a mature polypeptide.

5           In general, a NOVX -like variant that preserves NOVX-like function includes any variant in which residues at a particular position in the sequence have been substituted by other amino acids, and further include the possibility of inserting an additional residue or residues between two residues of the parent protein as well as the possibility of deleting one or more residues from the parent sequence. Any amino acid substitution, insertion, or deletion is encompassed by the  
10           invention. In favorable circumstances, the substitution is a conservative substitution as defined above.

One aspect of the invention pertains to isolated NOVX proteins, and biologically active portions thereof, or derivatives, fragments, analogs or homologs thereof. Also provided are polypeptide fragments suitable for use as immunogens to raise anti-NOVX antibodies. In one  
15           embodiment, native NOVX proteins can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, NOVX proteins are produced by recombinant DNA techniques. Alternative to recombinant expression, a NOVX protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

20           An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the NOVX protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of NOVX protein in which the protein is separated from cellular components of the  
25           cells from which it is isolated or recombinantly produced. In one embodiment, the language "substantially free of cellular material" includes preparations of NOVX protein having less than about 30% (by dry weight) of non-NOVX protein (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-NOVX protein, still more preferably less than about 10% of non-NOVX protein, and most preferably less than about 5% non-NOVX  
30           protein. When the NOVX protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, *i.e.*, culture medium

represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of NOVX protein in which the protein is separated from chemical precursors or other chemicals that are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of NOVX protein having less than about 30% (by dry weight) of chemical precursors or non-NOVX chemicals, more preferably less than about 20% chemical precursors or non-NOVX chemicals, still more preferably less than about 10% chemical precursors or non-NOVX chemicals, and most preferably less than about 5% chemical precursors or non-NOVX chemicals.

Biologically active portions of a NOVX protein include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequence of the NOVX protein, *e.g.*, the amino acid sequence shown in SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24 or 26 that include fewer amino acids than the full length NOVX proteins, and exhibit at least one activity of a NOVX protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the NOVX protein. A biologically active portion of a NOVX protein can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acids in length.

A biologically active portion of a NOVX protein of the present invention may contain at least one of the above-identified domains conserved between the NOVX proteins, *e.g.* TSR modules. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native NOVX protein.

In an embodiment, the NOVX protein has an amino acid sequence shown in SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24 or 26. In other embodiments, the NOVX protein is substantially homologous to SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24 or 26 and retains the functional activity of the protein of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24 or 26 yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail below. Accordingly, in another embodiment, the NOVX protein is a protein that comprises an amino acid sequence at least about 45% homologous to the amino acid sequence of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24 or 26 and retains the functional activity of the NOVX proteins of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24 or 26.

### Determining homology between two or more sequence

To determine the percent homology of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in either of the sequences being compared for optimal alignment between the sequences). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are homologous at that position (*i.e.*, as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity").

The nucleic acid sequence homology may be determined as the degree of identity between two sequences. The homology may be determined using computer programs known in the art, such as GAP software provided in the GCG program package. See, *Needleman and Wunsch 1970 J Mol Biol* 48: 443-453. Using GCG GAP software with the following settings for nucleic acid sequence comparison: GAP creation penalty of 5.0 and GAP extension penalty of 0.3, the coding region of the analogous nucleic acid sequences referred to above exhibits a degree of identity preferably of at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99%, with the CDS (encoding) part of the DNA sequence shown in SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23 or 25.

The term "sequence identity" refers to the degree to which two polynucleotide or polypeptide sequences are identical on a residue-by-residue basis over a particular region of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over that region of comparison, determining the number of positions at which the identical nucleic acid base (*e.g.*, A, T, C, G, U, or I, in the case of nucleic acids) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the region of comparison (*i.e.*, the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The term "substantial identity" as used herein denotes a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 80 percent sequence identity, preferably at least 85 percent identity and often 90 to 95 percent sequence identity, more usually at least 99 percent sequence identity as compared to a reference sequence over a comparison region. The term "percentage of positive residues" is calculated by comparing two optimally

aligned sequences over that region of comparison, determining the number of positions at which the identical and conservative amino acid substitutions, as defined above, occur in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the region of comparison (*i.e.*, the window size), and multiplying the result by 100 to yield the percentage of positive residues.

### **Chimeric and fusion proteins**

The invention also provides NOVX chimeric or fusion proteins. As used herein, a NOVX "chimeric protein" or "fusion protein" comprises a NOVX polypeptide operatively linked to a non-NOVX polypeptide. An "NOVX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to NOVX, whereas a "non-NOVX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein that is not substantially homologous to the NOVX protein, *e.g.*, a protein that is different from the NOVX protein and that is derived from the same or a different organism. Within a NOVX fusion protein the NOVX polypeptide can correspond to all or a portion of a NOVX protein. In one embodiment, a NOVX fusion protein comprises at least one biologically active portion of a NOVX protein. In another embodiment, a NOVX fusion protein comprises at least two biologically active portions of a NOVX protein. Within the fusion protein, the term "operatively linked" is intended to indicate that the NOVX polypeptide and the non-NOVX polypeptide are fused in-frame to each other. The non-NOVX polypeptide can be fused to the N-terminus or C-terminus of the NOVX polypeptide.

For example, in one embodiment a NOVX fusion protein comprises a NOVX polypeptide operably linked to the extracellular domain of a second protein. Such fusion proteins can be further utilized in screening assays for compounds that modulate NOVX activity (such assays are described in detail below).

In another embodiment, the fusion protein is a GST-NOVX fusion protein in which the NOVX sequences are fused to the C-terminus of the GST (*i.e.*, glutathione S-transferase) sequences. Such fusion proteins can facilitate the purification of recombinant NOVX.

In another embodiment, the fusion protein is a NOVX-immunoglobulin fusion protein in which the NOVX sequences comprising one or more domains are fused to sequences derived from a member of the immunoglobulin protein family. The NOVX-immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered

to a subject to inhibit an interaction between a NOVX ligand and a NOVX protein on the surface of a cell, to thereby suppress NOVX-mediated signal transduction *in vivo*. In one nonlimiting example, a contemplated NOVX ligand of the invention is the NOVX receptor. The NOVX-immunoglobulin fusion proteins can be used to affect the bioavailability of a NOVX cognate ligand. Inhibition of the NOVX ligand/NOVX interaction may be useful therapeutically for both the treatment of proliferative and differentiative disorders, *e.g.*, cancer as well as modulating (*e.g.*, promoting or inhibiting) cell survival, as well as acute and chronic inflammatory disorders and hyperplastic wound healing, *e.g.* hypertrophic scars and keloids. Moreover, the NOVX-immunoglobulin fusion proteins of the invention can be used as immunogens to produce anti-NOVX antibodies in a subject, to purify NOVX ligands, and in screening assays to identify molecules that inhibit the interaction of NOVX with a NOVX ligand.

A NOVX chimeric or fusion protein of the invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, *e.g.*, by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, Ausubel et al. (eds.) CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). A NOVX-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the NOVX protein.

#### **NOVX agonists and antagonists**

The present invention also pertains to variants of the NOVX proteins that function as either NOVX agonists (mimetics) or as NOVX antagonists. Variants of the NOVX protein can be generated by mutagenesis, *e.g.*, discrete point mutation or truncation of the NOVX protein. An agonist of the NOVX protein can retain substantially the same, or a subset of, the biological

activities of the naturally occurring form of the NOVX protein. An antagonist of the NOVX protein can inhibit one or more of the activities of the naturally occurring form of the NOVX protein by, for example, competitively binding to a downstream or upstream member of a cellular signaling cascade which includes the NOVX protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. In one embodiment, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the NOVX proteins.

10 Variants of the NOVX protein that function as either NOVX agonists (mimetics) or as NOVX antagonists can be identified by screening combinatorial libraries of mutants, *e.g.*, truncation mutants, of the NOVX protein for NOVX protein agonist or antagonist activity. In one embodiment, a variegated library of NOVX variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of NOVX variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential NOVX sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.*, for phage display) containing the set of NOVX sequences therein. There are a variety of methods which can be used to produce libraries of potential NOVX variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential NOVX sequences. Methods for synthesizing degenerate oligonucleotides are known in the art (see, *e.g.*, Narang (1983) *Tetrahedron* 39:3; Itakura *et al.* (1984) *Annu Rev Biochem* 53:323; Itakura *et al.* (1984) *Science* 198:1056; Ike *et al.* (1983) *Nucl Acid Res* 11:477.

### Polypeptide libraries

30 In addition, libraries of fragments of the NOVX protein coding sequence can be used to generate a variegated population of NOVX fragments for screening and subsequent selection of variants of a NOVX protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of a NOVX coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the



double stranded DNA, renaturing the DNA to form double stranded DNA that can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal and internal fragments of various sizes of the NOVX protein.

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of NOVX proteins. The most widely used techniques, which are amenable to high throughput analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique that enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify NOVX variants (Arkin and Yourvan (1992) PNAS 89:7811-7815; Delgrave *et al.* (1993) Protein Engineering 6:327-331).

### NOVX Antibodies

Also included in the invention are antibodies to NOVX proteins, or fragments of NOVX proteins. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin (Ig) molecules, i.e., molecules that contain an antigen binding site that specifically binds (immunoreacts with) an antigen. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain,  $F_{ab}$ ,  $F_{ab}'$  and  $F_{(ab)2}$  fragments, and an  $F_{ab}$  expression library. In general, an antibody molecule obtained from humans relates to any of the classes IgG, IgM, IgA, IgE and IgD, which differ from one another by the nature of the heavy chain present in the molecule. Certain classes have subclasses as well, such as IgG<sub>1</sub>, IgG<sub>2</sub>, and others. Furthermore, in humans, the light chain may be a kappa chain or a lambda chain. Reference herein to antibodies includes a reference to all such classes, subclasses and types of human antibody species.

An isolated NOVX-related protein of the invention may be intended to serve as an



Manual, Harlow E, and Lane D, 1988, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, incorporated herein by reference). Some of these antibodies are discussed below.

### **Polyclonal Antibodies**

5 For the production of polyclonal antibodies, various suitable host animals (e.g., rabbit, goat, mouse or other mammal) may be immunized by one or more injections with the native protein, a synthetic variant thereof, or a derivative of the foregoing. An appropriate immunogenic preparation can contain, for example, the naturally occurring immunogenic protein, a chemically synthesized polypeptide representing the immunogenic protein, or a  
10 recombinantly expressed immunogenic protein. Furthermore, the protein may be conjugated to a second protein known to be immunogenic in the mammal being immunized. Examples of such immunogenic proteins include but are not limited to keyhole limpet hemocyanin, serum albumin, bovine thyroglobulin, and soybean trypsin inhibitor. The preparation can further include an adjuvant. Various adjuvants used to increase the immunological response include, but are not  
15 limited to, Freund's (complete and incomplete), mineral gels (e.g., aluminum hydroxide), surface active substances (e.g., lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, dinitrophenol, etc.), adjuvants usable in humans such as Bacille Calmette-Guerin and Corynebacterium parvum, or similar immunostimulatory agents. Additional examples of adjuvants which can be employed include MPL-TDM adjuvant (monophosphoryl Lipid A,  
20 synthetic trehalose dicorynomycolate).

The polyclonal antibody molecules directed against the immunogenic protein can be isolated from the mammal (e.g., from the blood) and further purified by well known techniques, such as affinity chromatography using protein A or protein G, which provide primarily the IgG fraction of immune serum. Subsequently, or alternatively, the specific antigen which is the target  
25 of the immunoglobulin sought, or an epitope thereof, may be immobilized on a column to purify the immune specific antibody by immunoaffinity chromatography. Purification of immunoglobulins is discussed, for example, by D. Wilkinson (The Scientist, published by The Scientist, Inc., Philadelphia PA, Vol. 14, No. 8 (April 17, 2000), pp. 25-28).

### **Monoclonal Antibodies**

The term "monoclonal antibody" (MAb) or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one molecular species of

antibody molecule consisting of a unique light chain gene product and a unique heavy chain gene product. In particular, the complementarity determining regions (CDRs) of the monoclonal antibody are identical in all the molecules of the population. MAbs thus contain an antigen binding site capable of immunoreacting with a particular epitope of the antigen characterized by a unique binding affinity for it.

Monoclonal antibodies can be prepared using hybridoma methods, such as those described by Kohler and Milstein, Nature, 256:495 (1975). In a hybridoma method, a mouse, hamster, or other appropriate host animal, is typically immunized with an immunizing agent to elicit lymphocytes that produce or are capable of producing antibodies that will specifically bind to the immunizing agent. Alternatively, the lymphocytes can be immunized in vitro.

The immunizing agent will typically include the protein antigen, a fragment thereof or a fusion protein thereof. Generally, either peripheral blood lymphocytes are used if cells of human origin are desired, or spleen cells or lymph node cells are used if non-human mammalian sources are desired. The lymphocytes are then fused with an immortalized cell line using a suitable fusing agent, such as polyethylene glycol, to form a hybridoma cell (Goding, Monoclonal Antibodies: Principles and Practice, Academic Press, (1986) pp. 59-103). Immortalized cell lines are usually transformed mammalian cells, particularly myeloma cells of rodent, bovine and human origin. Usually, rat or mouse myeloma cell lines are employed. The hybridoma cells can be cultured in a suitable culture medium that preferably contains one or more substances that inhibit the growth or survival of the unfused, immortalized cells. For example, if the parental cells lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT or HPRT), the culture medium for the hybridomas typically will include hypoxanthine, aminopterin, and thymidine ("HAT medium"), which substances prevent the growth of HGPRT-deficient cells.

Preferred immortalized cell lines are those that fuse efficiently, support stable high level expression of antibody by the selected antibody-producing cells, and are sensitive to a medium such as HAT medium. More preferred immortalized cell lines are murine myeloma lines, which can be obtained, for instance, from the Salk Institute Cell Distribution Center, San Diego, California and the American Type Culture Collection, Manassas, Virginia. Human myeloma and mouse-human heteromyeloma cell lines also have been described for the production of human monoclonal antibodies (Kozbor, J. Immunol., 133:3001 (1984); Brodeur et al., Monoclonal Antibody Production Techniques and Applications, Marcel Dekker, Inc., New York, (1987) pp. 51-63).

The culture medium in which the hybridoma cells are cultured can then be assayed for the presence of monoclonal antibodies directed against the antigen. Preferably, the binding specificity of monoclonal antibodies produced by the hybridoma cells is determined by immunoprecipitation or by an in vitro binding assay, such as radioimmunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA). Such techniques and assays are known in the art. The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson and Pollard, Anal. Biochem., 107:220 (1980). Preferably, antibodies having a high degree of specificity and a high binding affinity for the target antigen are isolated.

After the desired hybridoma cells are identified, the clones can be subcloned by limiting dilution procedures and grown by standard methods. Suitable culture media for this purpose include, for example, Dulbecco's Modified Eagle's Medium and RPMI-1640 medium. Alternatively, the hybridoma cells can be grown *in vivo* as ascites in a mammal.

The monoclonal antibodies secreted by the subclones can be isolated or purified from the culture medium or ascites fluid by conventional immunoglobulin purification procedures such as, for example, protein A-Sepharose, hydroxylapatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

The monoclonal antibodies can also be made by recombinant DNA methods, such as those described in U.S. Patent No. 4,816,567. DNA encoding the monoclonal antibodies of the invention can be readily isolated and sequenced using conventional procedures (e.g., by using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of murine antibodies). The hybridoma cells of the invention serve as a preferred source of such DNA. Once isolated, the DNA can be placed into expression vectors, which are then transfected into host cells such as simian COS cells, Chinese hamster ovary (CHO) cells, or myeloma cells that do not otherwise produce immunoglobulin protein, to obtain the synthesis of monoclonal antibodies in the recombinant host cells. The DNA also can be modified, for example, by substituting the coding sequence for human heavy and light chain constant domains in place of the homologous murine sequences (U.S. Patent No. 4,816,567; Morrison, Nature 368, 812-13 (1994)) or by covalently joining to the immunoglobulin coding sequence all or part of the coding sequence for a non-immunoglobulin polypeptide. Such a non-immunoglobulin polypeptide can be substituted for the constant domains of an antibody of the invention, or can be

substituted for the variable domains of one antigen-combining site of an antibody of the invention to create a chimeric bivalent antibody.

### **Humanized Antibodies**

5           The antibodies directed against the protein antigens of the invention can further comprise humanized antibodies or human antibodies. These antibodies are suitable for administration to humans without engendering an immune response by the human against the administered immunoglobulin. Humanized forms of antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', F(ab')<sub>2</sub> or other antigen-  
10 binding subsequences of antibodies) that are principally comprised of the sequence of a human immunoglobulin, and contain minimal sequence derived from a non-human immunoglobulin. Humanization can be performed following the method of Winter and co-workers (Jones et al., Nature, 321:522-525 (1986); Riechmann et al., Nature, 332:323-327 (1988); Verhoeyen et al., Science, 239:1534-1536 (1988)), by substituting rodent CDRs or CDR sequences for the  
15 corresponding sequences of a human antibody. (See also U.S. Patent No. 5,225,539.) In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies can also comprise residues which are found neither in the recipient antibody nor in the imported CDR or framework sequences. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable  
20 domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the framework regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin (Jones et al., 1986; Riechmann et al., 1988; and Presta, Curr. Op. Struct. Biol.,  
25 2:593-596 (1992)).

### **Human Antibodies**

          Fully human antibodies relate to antibody molecules in which essentially the entire sequences of both the light chain and the heavy chain, including the CDRs, arise from human  
30 genes. Such antibodies are termed "human antibodies", or "fully human antibodies" herein. Human monoclonal antibodies can be prepared by the trioma technique; the human B-cell hybridoma technique (see Kozbor, et al., 1983 Immunol Today 4: 72) and the EBV hybridoma



technique to produce human monoclonal antibodies (see Cole, et al., 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96). Human monoclonal antibodies may be utilized in the practice of the present invention and may be produced by using human hybridomas (see Cote, et al., 1983. Proc Natl Acad Sci USA 80: 2026-2030) or by transforming human B-cells with Epstein Barr Virus in vitro (see Cole, et al., 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96).

In addition, human antibodies can also be produced using additional techniques, including phage display libraries (Hoogenboom and Winter, *J. Mol. Biol.*, 227:381 (1991); Marks et al., *J. Mol. Biol.*, 222:581 (1991)). Similarly, human antibodies can be made by introducing human immunoglobulin loci into transgenic animals, e.g., mice in which the endogenous immunoglobulin genes have been partially or completely inactivated. Upon challenge, human antibody production is observed, which closely resembles that seen in humans in all respects, including gene rearrangement, assembly, and antibody repertoire. This approach is described, for example, in U.S. Patent Nos. 5,545,807; 5,545,806; 5,569,825; 5,625,126; 5,633,425; 5,661,016, and in Marks et al. (*Bio/Technology* 10, 779-783 (1992)); Lonberg et al. (*Nature* 368 856-859 (1994)); Morrison (*Nature* 368, 812-13 (1994)); Fishwild et al., (*Nature Biotechnology* 14, 845-51 (1996)); Neuberger (*Nature Biotechnology* 14, 826 (1996)); and Lonberg and Huszar (*Intern. Rev. Immunol.* 13 65-93 (1995)).

Human antibodies may additionally be produced using transgenic nonhuman animals which are modified so as to produce fully human antibodies rather than the animal's endogenous antibodies in response to challenge by an antigen. (See PCT publication WO94/02602). The endogenous genes encoding the heavy and light immunoglobulin chains in the nonhuman host have been incapacitated, and active loci encoding human heavy and light chain immunoglobulins are inserted into the host's genome. The human genes are incorporated, for example, using yeast artificial chromosomes containing the requisite human DNA segments. An animal which provides all the desired modifications is then obtained as progeny by crossbreeding intermediate transgenic animals containing fewer than the full complement of the modifications. The preferred embodiment of such a nonhuman animal is a mouse, and is termed the Xenomouse<sup>TM</sup> as disclosed in PCT publications WO 96/33735 and WO 96/34096. This animal produces B cells which secrete fully human immunoglobulins. The antibodies can be obtained directly from the animal after immunization with an immunogen of interest, as, for example, a preparation of a polyclonal antibody, or alternatively from immortalized B cells derived from the animal, such as

hybridomas producing monoclonal antibodies. Additionally, the genes encoding the immunoglobulins with human variable regions can be recovered and expressed to obtain the antibodies directly, or can be further modified to obtain analogs of antibodies such as, for example, single chain Fv molecules.

5           An example of a method of producing a nonhuman host, exemplified as a mouse, lacking expression of an endogenous immunoglobulin heavy chain is disclosed in U.S. Patent No. 5,939,598. It can be obtained by a method including deleting the J segment genes from at least one endogenous heavy chain locus in an embryonic stem cell to prevent rearrangement of the locus and to prevent formation of a transcript of a rearranged immunoglobulin heavy chain locus,  
10       the deletion being effected by a targeting vector containing a gene encoding a selectable marker; and producing from the embryonic stem cell a transgenic mouse whose somatic and germ cells contain the gene encoding the selectable marker.

          A method for producing an antibody of interest, such as a human antibody, is disclosed in U.S. Patent No. 5,916,771. It includes introducing an expression vector that contains a  
15       nucleotide sequence encoding a heavy chain into one mammalian host cell in culture, introducing an expression vector containing a nucleotide sequence encoding a light chain into another mammalian host cell, and fusing the two cells to form a hybrid cell. The hybrid cell expresses an antibody containing the heavy chain and the light chain.

          In a further improvement on this procedure, a method for identifying a clinically relevant  
20       epitope on an immunogen, and a correlative method for selecting an antibody that binds immunospecifically to the relevant epitope with high affinity, are disclosed in PCT publication WO 99/53049.

#### **F<sub>ab</sub> Fragments and Single Chain Antibodies**

25           According to the invention, techniques can be adapted for the production of single-chain antibodies specific to an antigenic protein of the invention (see e.g., U.S. Patent No. 4,946,778). In addition, methods can be adapted for the construction of F<sub>ab</sub> expression libraries (see e.g., Huse, et al., 1989 Science 246: 1275-1281) to allow rapid and effective identification of monoclonal F<sub>ab</sub> fragments with the desired specificity for a protein or derivatives, fragments,  
30       analogs or homologs thereof. Antibody fragments that contain the idiotypes to a protein antigen may be produced by techniques known in the art including, but not limited to: (i) an F<sub>(ab')<sub>2</sub></sub> fragment produced by pepsin digestion of an antibody molecule; (ii) an F<sub>ab</sub> fragment generated by

reducing the disulfide bridges of an  $F_{(ab)2}$  fragment; (iii) an  $F_{ab}$  fragment generated by the treatment of the antibody molecule with papain and a reducing agent and (iv)  $F_v$  fragments.

### Bispecific Antibodies

5 Bispecific antibodies are monoclonal, preferably human or humanized, antibodies that have binding specificities for at least two different antigens. In the present case, one of the binding specificities is for an antigenic protein of the invention. The second binding target is any other antigen, and advantageously is a cell-surface protein or receptor or receptor subunit.

Methods for making bispecific antibodies are known in the art. Traditionally, the  
10 recombinant production of bispecific antibodies is based on the co-expression of two immunoglobulin heavy-chain/light-chain pairs, where the two heavy chains have different specificities (Milstein and Cuello, *Nature*, 305:537-539 (1983)). Because of the random assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas) produce a potential mixture of ten different antibody molecules, of which only one has the correct  
15 bispecific structure. The purification of the correct molecule is usually accomplished by affinity chromatography steps. Similar procedures are disclosed in WO 93/08829, published 13 May 1993, and in Traunecker *et al.*, 1991 *EMBO J.*, 10:3655-3659.

Antibody variable domains with the desired binding specificities (antibody-antigen combining sites) can be fused to immunoglobulin constant domain sequences. The fusion  
20 preferably is with an immunoglobulin heavy-chain constant domain, comprising at least part of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1) containing the site necessary for light-chain binding present in at least one of the fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the immunoglobulin light chain, are inserted into separate expression vectors, and are co-transfected into a suitable  
25 host organism. For further details of generating bispecific antibodies see, for example, Suresh *et al.*, *Methods in Enzymology*, 121:210 (1986).

According to another approach described in WO 96/27011, the interface between a pair of antibody molecules can be engineered to maximize the percentage of heterodimers which are recovered from recombinant cell culture. The preferred interface comprises at least a part of the  
30 CH3 region of an antibody constant domain. In this method, one or more small amino acid side chains from the interface of the first antibody molecule are replaced with larger side chains (e.g. tyrosine or tryptophan). Compensatory "cavities" of identical or similar size to the large side chain(s) are created on the interface of the second antibody molecule by replacing large amino

acid side chains with smaller ones (e.g. alanine or threonine). This provides a mechanism for increasing the yield of the heterodimer over other unwanted end-products such as homodimers.

Bispecific antibodies can be prepared as full length antibodies or antibody fragments (e.g.  $F(ab')_2$  bispecific antibodies). Techniques for generating bispecific antibodies from antibody fragments have been described in the literature. For example, bispecific antibodies can be prepared using chemical linkage. Brennan et al., Science 229:81 (1985) describe a procedure wherein intact antibodies are proteolytically cleaved to generate  $F(ab')_2$  fragments. These fragments are reduced in the presence of the dithiol complexing agent sodium arsenite to stabilize vicinal dithiols and prevent intermolecular disulfide formation. The  $Fab'$  fragments generated are then converted to thionitrobenzoate (TNB) derivatives. One of the  $Fab'$ -TNB derivatives is then reconverted to the  $Fab'$ -thiol by reduction with mercaptoethylamine and is mixed with an equimolar amount of the other  $Fab'$ -TNB derivative to form the bispecific antibody. The bispecific antibodies produced can be used as agents for the selective immobilization of enzymes.

Additionally,  $Fab'$  fragments can be directly recovered from *E. coli* and chemically coupled to form bispecific antibodies. Shalaby et al., J. Exp. Med. 175:217-225 (1992) describe the production of a fully humanized bispecific antibody  $F(ab')_2$  molecule. Each  $Fab'$  fragment was separately secreted from *E. coli* and subjected to directed chemical coupling in vitro to form the bispecific antibody. The bispecific antibody thus formed was able to bind to cells overexpressing the ErbB2 receptor and normal human T cells, as well as trigger the lytic activity of human cytotoxic lymphocytes against human breast tumor targets.

Various techniques for making and isolating bispecific antibody fragments directly from recombinant cell culture have also been described. For example, bispecific antibodies have been produced using leucine zippers. Kostelny et al., J. Immunol. 148(5):1547-1553 (1992). The leucine zipper peptides from the Fos and Jun proteins were linked to the  $Fab'$  portions of two different antibodies by gene fusion. The antibody homodimers were reduced at the hinge region to form monomers and then re-oxidized to form the antibody heterodimers. This method can also be utilized for the production of antibody homodimers. The "diabody" technology described by Hollinger et al., Proc. Natl. Acad. Sci. USA 90:6444-6448 (1993) has provided an alternative mechanism for making bispecific antibody fragments. The fragments comprise a heavy-chain variable domain ( $V_H$ ) connected to a light-chain variable domain ( $V_L$ ) by a linker which is too short to allow pairing between the two domains on the same chain. Accordingly, the  $V_H$  and  $V_L$

domains of one fragment are forced to pair with the complementary  $V_L$  and  $V_H$  domains of another fragment, thereby forming two antigen-binding sites. Another strategy for making bispecific antibody fragments by the use of single-chain Fv (sFv) dimers has also been reported. See, Gruber et al., J. Immunol. 152:5368 (1994).

5           Antibodies with more than two valencies are contemplated. For example, trispecific antibodies can be prepared. Tutt et al., J. Immunol. 147:60 (1991).

Exemplary bispecific antibodies can bind to two different epitopes, at least one of which originates in the protein antigen of the invention. Alternatively, an anti-antigenic arm of an immunoglobulin molecule can be combined with an arm which binds to a triggering molecule on  
10 a leukocyte such as a T-cell receptor molecule (e.g. CD2, CD3, CD28, or B7), or Fc receptors for IgG (FcγR), such as FcγRI (CD64), FcγRII (CD32) and FcγRIII (CD16) so as to focus cellular defense mechanisms to the cell expressing the particular antigen. Bispecific antibodies can also be used to direct cytotoxic agents to cells which express a particular antigen. These antibodies possess an antigen-binding arm and an arm which binds a cytotoxic agent or a radionuclide  
15 chelator, such as EOTUBE, DPTA, DOTA, or TETA. Another bispecific antibody of interest binds the protein antigen described herein and further binds tissue factor (TF).

### **Heteroconjugate Antibodies**

Heteroconjugate antibodies are also within the scope of the present invention.

20 Heteroconjugate antibodies are composed of two covalently joined antibodies. Such antibodies have, for example, been proposed to target immune system cells to unwanted cells (U.S. Patent No. 4,676,980), and for treatment of HIV infection (WO 91/00360; WO 92/200373; EP 03089). It is contemplated that the antibodies can be prepared in vitro using known methods in synthetic protein chemistry, including those involving crosslinking agents. For example, immunotoxins  
25 can be constructed using a disulfide exchange reaction or by forming a thioether bond. Examples of suitable reagents for this purpose include iminothiolate and methyl-4-mercaptobutyrimidate and those disclosed, for example, in U.S. Patent No. 4,676,980.

### **Effector Function Engineering**

30           It can be desirable to modify the antibody of the invention with respect to effector function, so as to enhance, e.g., the effectiveness of the antibody in treating cancer. For example, cysteine residue(s) can be introduced into the Fc region, thereby allowing interchain disulfide bond formation in this region. The homodimeric antibody thus generated can have improved

internalization capability and/or increased complement-mediated cell killing and antibody-dependent cellular cytotoxicity (ADCC). See Caron et al., *J. Exp Med.*, 176: 1191-1195 (1992) and Shopes, *J. Immunol.*, 148: 2918-2922 (1992). Homodimeric antibodies with enhanced anti-tumor activity can also be prepared using heterobifunctional cross-linkers as described in Wolff et al. *Cancer Research*, 53: 2560-2565 (1993). Alternatively, an antibody can be engineered that has dual Fc regions and can thereby have enhanced complement lysis and ADCC capabilities. See Stevenson et al., *Anti-Cancer Drug Design*, 3: 219-230 (1989).

### Immunoconjugates

The invention also pertains to immunoconjugates comprising an antibody conjugated to a cytotoxic agent such as a chemotherapeutic agent, toxin (e.g., an enzymatically active toxin of bacterial, fungal, plant, or animal origin, or fragments thereof), or a radioactive isotope (i.e., a radioconjugate).

Chemotherapeutic agents useful in the generation of such immunoconjugates have been described above. Enzymatically active toxins and fragments thereof that can be used include diphtheria A chain, nonbinding active fragments of diphtheria toxin, exotoxin A chain (from *Pseudomonas aeruginosa*), ricin A chain, abrin A chain, modeccin A chain, alpha-sarcin, Aleurites fordii proteins, dianthin proteins, *Phytolacca americana* proteins (PAPI, PAPII, and PAP-S), momordica charantia inhibitor, curcin, crotin, sapaonaria officinalis inhibitor, gelonin, mitogellin, restrictocin, phenomycin, enomycin, and the tricothecenes. A variety of radionuclides are available for the production of radioconjugated antibodies. Examples include  $^{212}\text{Bi}$ ,  $^{131}\text{I}$ ,  $^{131}\text{In}$ ,  $^{90}\text{Y}$ , and  $^{186}\text{Re}$ .

Conjugates of the antibody and cytotoxic agent are made using a variety of bifunctional protein-coupling agents such as N-succinimidyl-3-(2-pyridyldithiol) propionate (SPDP), iminothiolane (IT), bifunctional derivatives of imidoesters (such as dimethyl adipimidate HCL), active esters (such as disuccinimidyl suberate), aldehydes (such as glutaraldehyde), bis-azido compounds (such as bis (p-azidobenzoyl) hexanediamine), bis-diazonium derivatives (such as bis-(p-diazoniumbenzoyl)-ethylenediamine), diisocyanates (such as tolyene 2,6-diisocyanate), and bis-active fluorine compounds (such as 1,5-difluoro-2,4-dinitrobenzene). For example, a ricin immunotoxin can be prepared as described in Vitetta et al., *Science*, 238: 1098 (1987). Carbon-14-labeled 1-isothiocyanatobenzyl-3-methyldiethylene triaminepentaacetic acid (MX-



DTPA) is an exemplary chelating agent for conjugation of radionucleotide to the antibody. See WO94/11026.

5 In another embodiment, the antibody can be conjugated to a "receptor" (such streptavidin) for utilization in tumor pretargeting wherein the antibody-receptor conjugate is administered to the patient, followed by removal of unbound conjugate from the circulation using a clearing agent and then administration of a "ligand" (e.g., avidin) that is in turn conjugated to a cytotoxic agent.

### NOVX Recombinant Expression Vectors and Host Cells

10 Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding a NOVX protein, or derivatives, fragments, analogs or homologs thereof. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be  
15 ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are  
20 replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively-linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the  
25 invention is intended to include such other forms of expression vectors, such as viral vectors (e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the  
30 recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably-linked" is intended to mean

that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner that allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell).

5 The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (*e.g.*, NOVX proteins, mutant forms of NOVX proteins, fusion proteins, etc.).

15 The recombinant expression vectors of the invention can be designed for expression of NOVX proteins in prokaryotic or eukaryotic cells. For example, NOVX proteins can be expressed in bacterial cells such as *Escherichia coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

20 Expression of proteins in prokaryotes is most often carried out in *Escherichia coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: (i) to increase expression of recombinant protein; (ii) to increase the solubility of the recombinant protein; and (iii) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa,

thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson, 1988. *Gene* 67: 31-40), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) that fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

5        Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amrann *et al.*, (1988) *Gene* 69:301-315) and pET 11d (Studier *et al.*, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 60-89).

10        One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein. *See, e.g.*, Gottesman, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 119-128. Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (*see, e.g.*, Wada, *et al.*, 1992. *Nucl. Acids Res.* 20: 2111-2118). Such alteration of nucleic acid sequences of the invention can be  
15        carried out by standard DNA synthesis techniques.

20        In another embodiment, the NOVX expression vector is a yeast expression vector. Examples of vectors for expression in yeast *Saccharomyces cerevisiae* include pYepSec1 (Baldari, *et al.*, 1987. *EMBO J.* 6: 229-234), pMFa (Kurjan and Herskowitz, 1982. *Cell* 30: 933-943), pJRY88 (Schultz *et al.*, 1987. *Gene* 54: 113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), and picZ (InVitrogen Corp, San Diego, Calif.).

25        Alternatively, NOVX can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (*e.g.*, SF9 cells) include the pAc series (Smith, *et al.*, 1983. *Mol. Cell. Biol.* 3: 2156-2165) and the pVL series (Lucklow and Summers, 1989. *Virology* 170: 31-39).

30        In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, 1987. *Nature* 329: 840) and pMT2PC (Kaufman, *et al.*, 1987. *EMBO J.* 6: 187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, adenovirus 2, cytomegalovirus, and simian virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see, *e.g.*, Chapters 16 and 17 of Sambrook, *et*

al., MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert, *et al.*, 1987. *Genes Dev.* 1: 268-277), lymphoid-specific promoters (Calame and Eaton, 1988. *Adv. Immunol.* 43: 235-275), in particular promoters of T cell receptors (Winoto and Baltimore, 1989. *EMBO J.* 8: 729-733) and immunoglobulins (Banerji, *et al.*, 1983. *Cell* 33: 729-740; Queen and Baltimore, 1983. *Cell* 33: 741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and Ruddle, 1989. *Proc. Natl. Acad. Sci. USA* 86: 5473-5477), pancreas-specific promoters (Edlund, *et al.*, 1985. *Science* 230: 912-916), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Pat. No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, e.g., the murine hox promoters (Kessel and Gruss, 1990. *Science* 249: 374-379) and the  $\alpha$ -fetoprotein promoter (Campes and Tilghman, 1989. *Genes Dev.* 3: 537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively-linked to a regulatory sequence in a manner that allows for expression (by transcription of the DNA molecule) of an RNA molecule that is antisense to NOVX mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen that direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen that direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see, e.g., Weintraub, *et al.*, "Antisense RNA as a molecular tool for genetic analysis," *Reviews-Trends in Genetics*, Vol. 1(1) 1986.

Another aspect of the invention pertains to host cells into which a recombinant

expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, NOVX protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as human, Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Various selectable markers include those that confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding NOVX or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) NOVX protein. Accordingly, the invention further provides methods for producing NOVX protein using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression

vector encoding NOVX protein has been introduced) in a suitable medium such that NOVX protein is produced. In another embodiment, the method further comprises isolating NOVX protein from the medium or the host cell.

## 5     **Transgenic NOVX Animals**

10     The host cells of the invention can also be used to produce non-human transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which NOVX protein-coding sequences have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous NOVX sequences have been introduced into their genome or homologous recombinant animals in which endogenous NOVX sequences have been altered. Such animals are useful for studying the function and/or activity of NOVX protein and for identifying and/or evaluating modulators of NOVX protein activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, etc. A transgene is exogenous DNA that is integrated into the genome of a cell from which a transgenic animal develops and that remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, a

15     "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous NOVX gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, *e.g.*, an embryonic cell of the animal, prior to development of the animal.

20     A transgenic animal of the invention can be created by introducing NOVX-encoding nucleic acid into the male pronuclei of a fertilized oocyte (*e.g.*, by microinjection, retroviral infection) and allowing the oocyte to develop in a pseudopregnant female foster animal. Sequences including SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23 or 25 can be introduced as a transgene into the genome of a non-human animal. Alternatively, a non-human homologue of the human NOVX gene, such as a mouse NOVX gene, can be isolated based on hybridization

25     to the human NOVX cDNA (described further *supra*) and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be

30



operably-linked to the NOVX transgene to direct expression of NOVX protein to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866; 4,870,009; and 4,873,191; and Hogan, 1986. In:

5 MANIPULATING THE MOUSE EMBRYO, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the NOVX transgene in its genome and/or expression of NOVX mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals  
10 carrying a transgene-encoding NOVX protein can further be bred to other transgenic animals carrying other transgenes.

To create a homologous recombinant animal, a vector is prepared which contains at least a portion of a NOVX gene into which a deletion, addition or substitution has been introduced to thereby alter, *e.g.*, functionally disrupt, the NOVX gene. The NOVX gene can be a human gene  
15 (*e.g.*, the DNA of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23 or 25), but more preferably, is a non-human homologue of a human NOVX gene. For example, a mouse homologue of human NOVX gene of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23 or 25 can be used to construct a homologous recombination vector suitable for altering an endogenous NOVX gene in the mouse genome. In one embodiment, the vector is designed such that, upon  
20 homologous recombination, the endogenous NOVX gene is functionally disrupted (*i.e.*, no longer encodes a functional protein; also referred to as a "knock out" vector).

Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous NOVX gene is mutated or otherwise altered but still encodes functional protein (*e.g.*, the upstream regulatory region can be altered to thereby alter the expression of the endogenous  
25 NOVX protein). In the homologous recombination vector, the altered portion of the NOVX gene is flanked at its 5'- and 3'-termini by additional nucleic acid of the NOVX gene to allow for homologous recombination to occur between the exogenous NOVX gene carried by the vector and an endogenous NOVX gene in an embryonic stem cell. The additional flanking NOVX nucleic acid is of sufficient length for successful homologous recombination with the  
30 endogenous gene. Typically, several kilobases of flanking DNA (both at the 5'- and 3'-termini) are included in the vector. *See, e.g.*, Thomas, *et al.*, 1987. *Cell* 51: 503 for a description of homologous recombination vectors. The vector is then introduced into an embryonic stem cell

line (e.g., by electroporation) and cells in which the introduced NOVX gene has homologously-recombined with the endogenous NOVX gene are selected. *See, e.g., Li, et al., 1992. Cell* 69: 915.

5 The selected cells are then injected into a blastocyst of an animal (e.g., a mouse) to form aggregation chimeras. *See, e.g., Bradley, 1987. In: TERATOCARCINOMAS AND EMBRYONIC STEM CELLS: A PRACTICAL APPROACH, Robertson, ed. IRL, Oxford, pp. 113-152.* A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously-recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously-recombined  
10 DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley, 1991. *Curr. Opin. Biotechnol.* 2: 823-829; PCT International Publication Nos.: WO 90/11354; WO 91/01140; WO 92/0968; and WO 93/04169.

15 In another embodiment, transgenic non-humans animals can be produced that contain selected systems that allow for regulated expression of the transgene. One example of such a system is the cre/loxP recombinase system of bacteriophage P1. For a description of the cre/loxP recombinase system, *See, e.g., Lakso, et al., 1992. Proc. Natl. Acad. Sci. USA* 89: 6232-6236. Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae*. *See, O'Gorman, et al., 1991. Science* 251:1351-1355. If a cre/loxP recombinase  
20 system is used to regulate expression of the transgene, animals containing transgenes encoding both the Cre recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

25 Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, *et al., 1997. Nature* 385: 810-813. In brief, a cell (e.g., a somatic cell) from the transgenic animal can be isolated and induced to exit the growth cycle and enter G<sub>0</sub> phase. The quiescent cell can then be fused, e.g., through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell  
30 is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyte and then transferred to pseudopregnant female foster animal. The offspring borne of

this female foster animal will be a clone of the animal from which the cell (*e.g.*, the somatic cell) is isolated.

### Pharmaceutical Compositions

5           The NOVX nucleic acid molecules, NOVX proteins, and anti-NOVX antibodies (also referred to herein as "active compounds") of the invention, and derivatives, fragments, analogs and homologs thereof, can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein, "pharmaceutically  
10 acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Suitable carriers are described in the most recent edition of Remington's Pharmaceutical Sciences, a standard reference text in the field, which is incorporated herein by reference. Preferred examples of such carriers or diluents  
15 include, but are not limited to, water, saline, finger's solutions, dextrose solution, and 5% human serum albumin. Liposomes and non-aqueous vehicles such as fixed oils may also be used. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be  
20 incorporated into the compositions.

          The antibodies disclosed herein can also be formulated as immunoliposomes. Liposomes containing the antibody are prepared by methods known in the art, such as described in Epstein et al., Proc. Natl. Acad. Sci. USA, 82: 3688 (1985); Hwang et al., Proc. Natl. Acad. Sci. USA, 77: 4030 (1980); and U.S. Pat. Nos. 4,485,045 and 4,544,545. Liposomes with  
25 enhanced circulation time are disclosed in U.S. Patent No. 5,013,556.

          Particularly useful liposomes can be generated by the reverse-phase evaporation method with a lipid composition comprising phosphatidylcholine, cholesterol, and PEG-derivatized phosphatidylethanolamine (PEG-PE). Liposomes are extruded through filters of defined pore size to yield liposomes with the desired diameter. Fab' fragments of the antibody of the present  
30 invention can be conjugated to the liposomes as described in Martin et al., J. Biol. Chem., 257: 286-288 (1982) via a disulfide-interchange reaction. A chemotherapeutic agent (such as

Doxorubicin) is optionally contained within the liposome. See Gabizon et al., J. National Cancer Inst., 81(19): 1484 (1989).

5 A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (*i.e.*, topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; 10 antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid (EDTA); buffers such as acetates, citrates or phosphates, and agents for the adjustment of tonicity such as sodium chloride or dextrose. The pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of 15 glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, N.J.) or 20 phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringeability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and 25 liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, 30 it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable

compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

5 Sterile injectable solutions can be prepared by incorporating the active compound (*e.g.*, a NOVX protein or anti-NOVX antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle that contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, methods of preparation are vacuum drying and freeze-drying that yields a powder of  
10 the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of  
15 tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a  
20 disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

25 For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, *e.g.*, a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated  
30 are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or

suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (*see, e.g.*, U.S. Patent No. 5,328,470) or by stereotactic injection (*see, e.g.*, Chen, *et al.*, 1994. *Proc. Natl. Acad. Sci. USA* 91: 3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant



cells, *e.g.*, retroviral vectors, the pharmaceutical preparation can include one or more cells that produce the gene delivery system.

Antibodies specifically binding a protein of the invention, as well as other molecules identified by the screening assays disclosed herein, can be administered for the treatment of various disorders in the form of pharmaceutical compositions. Principles and considerations involved in preparing such compositions, as well as guidance in the choice of components are provided, for example, in Remington : The Science And Practice Of Pharmacy 19th ed. (Alfonso R. Gennaro, et al., editors) Mack Pub. Co., Easton, Pa. : 1995; Drug Absorption Enhancement : Concepts, Possibilities, Limitations, And Trends, Harwood Academic Publishers, Langhorne, Pa., 1994; and Peptide And Protein Drug Delivery (Advances In Parenteral Sciences, Vol. 4), 1991, M. Dekker, New York. If the antigenic protein is intracellular and whole antibodies are used as inhibitors, internalizing antibodies are preferred. However, liposomes can also be used to deliver the antibody, or an antibody fragment, into cells. Where antibody fragments are used, the smallest inhibitory fragment that specifically binds to the binding domain of the target protein is preferred. For example, based upon the variable-region sequences of an antibody, peptide molecules can be designed that retain the ability to bind the target protein sequence. Such peptides can be synthesized chemically and/or produced by recombinant DNA technology. See, *e.g.*, Marasco *et al.*, 1993 *Proc. Natl. Acad. Sci. USA*, 90: 7889-7893. The formulation herein can also contain more than one active compound as necessary for the particular indication being treated, preferably those with complementary activities that do not adversely affect each other. Alternatively, or in addition, the composition can comprise an agent that enhances its function, such as, for example, a cytotoxic agent, cytokine, chemotherapeutic agent, or growth-inhibitory agent. Such molecules are suitably present in combination in amounts that are effective for the purpose intended. The active ingredients can also be entrapped in microcapsules prepared, for example, by coacervation techniques or by interfacial polymerization, for example, hydroxymethylcellulose or gelatin-microcapsules and poly-(methylmethacrylate) microcapsules, respectively, in colloidal drug delivery systems (for example, liposomes, albumin microspheres, microemulsions, nano-particles, and nanocapsules) or in macroemulsions.

The formulations to be used for *iv vivo* administration must be sterile. This is readily accomplished by filtration through sterile filtration membranes.

Sustained-release preparations can be prepared. Suitable examples of sustained-release preparations include semipermeable matrices of solid hydrophobic polymers containing the

antibody, which matrices are in the form of shaped articles, e.g., films, or microcapsules. Examples of sustained-release matrices include polyesters, hydrogels (for example, poly(2-hydroxyethyl-methacrylate), or poly(vinylalcohol)), polylactides (U.S. Pat. No. 3,773,919), copolymers of L-glutamic acid and  $\gamma$  ethyl-L-glutamate, non-degradable ethylene-vinyl acetate, 5 degradable lactic acid-glycolic acid copolymers such as the LUPRON DEPOT™ (injectable microspheres composed of lactic acid-glycolic acid copolymer and leuprolide acetate), and poly-D-(-)-3-hydroxybutyric acid. While polymers such as ethylene-vinyl acetate and lactic acid-glycolic acid enable release of molecules for over 100 days, certain hydrogels release proteins for shorter time periods.

10 The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

### Screening and Detection Methods

15 The isolated nucleic acid molecules of the invention can be used to express NOVX protein (e.g., via a recombinant expression vector in a host cell in gene therapy applications), to detect NOVX mRNA (e.g., in a biological sample) or a genetic lesion in a NOVX gene, and to modulate NOVX activity, as described further, below. In addition, the NOVX proteins can be used to screen drugs or compounds that modulate the NOVX protein activity or expression as well as to treat disorders characterized by insufficient or excessive production of NOVX protein 20 or production of NOVX protein forms that have decreased or aberrant activity compared to NOVX wild-type protein. In addition, the anti-NOVX antibodies of the invention can be used to detect and isolate NOVX proteins and modulate NOVX activity. For example, NOVX activity includes growth and differentiation, antibody production, and tumor growth.

25 The invention further pertains to novel agents identified by the screening assays described herein and uses thereof for treatments as described, *supra*.

### Screening Assays

30 The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, *i.e.*, candidate or test compounds or agents (*e.g.*, peptides, peptidomimetics, small molecules or other drugs) that bind to NOVX proteins or have a stimulatory or inhibitory effect on, *e.g.*, NOVX protein expression or NOVX protein activity. The invention also includes compounds identified in the screening assays described herein.

10 In one embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of the membrane-bound form of a NOVX protein or polypeptide or biologically-active portion thereof. The test compounds of the invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the "one-bead one-compound" library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds. See, e.g., Lam, 1997. *Anticancer Drug Design* 12: 145.

15 A "small molecule" as used herein, is meant to refer to a composition that has a molecular weight of less than about 5 kD and most preferably less than about 4 kD. Small molecules can be, e.g., nucleic acids, peptides, polypeptides, peptidomimetics, carbohydrates, lipids or other organic or inorganic molecules. Libraries of chemical and/or biological mixtures, such as fungal, bacterial, or algal extracts, are known in the art and can be screened with any of the assays of the invention.

20 Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt, *et al.*, 1993. *Proc. Natl. Acad. Sci. U.S.A.* 90: 6909; Erb, *et al.*, 1994. *Proc. Natl. Acad. Sci. U.S.A.* 91: 11422; Zuckermann, *et al.*, 1994. *J. Med. Chem.* 37: 2678; Cho, *et al.*, 1993. *Science* 261: 1303; Carrell, *et al.*, 1994. *Angew. Chem. Int. Ed. Engl.* 33: 2059; Carell, *et al.*, 1994. *Angew. Chem. Int. Ed. Engl.* 33: 2061; and Gallop, *et al.*, 1994. *J. Med. Chem.* 37: 1233.

25 Libraries of compounds may be presented in solution (e.g., Houghten, 1992. *Biotechniques* 13: 412-421), or on beads (Lam, 1991. *Nature* 354: 82-84), on chips (Fodor, 1993. *Nature* 364: 555-556), bacteria (Ladner, U.S. Patent No. 5,223,409), spores (Ladner, U.S. Patent 5,233,409), plasmids (Cull, *et al.*, 1992. *Proc. Natl. Acad. Sci. USA* 89: 1865-1869) or on phage (Scott and Smith, 1990. *Science* 249: 386-390; Devlin, 1990. *Science* 249: 404-406; Cwirla, *et al.*, 1990. *Proc. Natl. Acad. Sci. U.S.A.* 87: 6378-6382; Felici, 1991. *J. Mol. Biol.* 222: 301-310; Ladner, U.S. Patent No. 5,233,409.).

30 In one embodiment, an assay is a cell-based assay in which a cell which expresses a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface is contacted with a test compound and the ability of the test compound to bind to a

NOVX protein determined. The cell, for example, can be of mammalian origin or a yeast cell. Determining the ability of the test compound to bind to the NOVX protein can be accomplished, for example, by coupling the test compound with a radioisotope or enzymatic label such that binding of the test compound to the NOVX protein or biologically-active portion thereof can be determined by detecting the labeled compound in a complex. For example, test compounds can be labeled with  $^{125}\text{I}$ ,  $^{35}\text{S}$ ,  $^{14}\text{C}$ , or  $^3\text{H}$ , either directly or indirectly, and the radioisotope detected by direct counting of radioemission or by scintillation counting. Alternatively, test compounds can be enzymatically-labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product. In one embodiment, the assay comprises contacting a cell which expresses a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface with a known compound which binds NOVX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a NOVX protein, wherein determining the ability of the test compound to interact with a NOVX protein comprises determining the ability of the test compound to preferentially bind to NOVX protein or a biologically-active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface with a test compound and determining the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the NOVX protein or biologically-active portion thereof. Determining the ability of the test compound to modulate the activity of NOVX or a biologically-active portion thereof can be accomplished, for example, by determining the ability of the NOVX protein to bind to or interact with a NOVX target molecule. As used herein, a "target molecule" is a molecule with which a NOVX protein binds or interacts in nature, for example, a molecule on the surface of a cell which expresses a NOVX interacting protein, a molecule on the surface of a second cell, a molecule in the extracellular milieu, a molecule associated with the internal surface of a cell membrane or a cytoplasmic molecule. A NOVX target molecule can be a non-NOVX molecule or a NOVX protein or polypeptide of the invention. In one embodiment, a NOVX target molecule is a component of a signal transduction pathway that facilitates transduction of an extracellular signal (*e.g.* a signal generated by binding of a compound to a membrane-bound NOVX molecule) through the cell membrane and into the

cell. The target, for example, can be a second intercellular protein that has catalytic activity or a protein that facilitates the association of downstream signaling molecules with NOVX.

Determining the ability of the NOVX protein to bind to or interact with a NOVX target molecule can be accomplished by one of the methods described above for determining direct  
5 binding. In one embodiment, determining the ability of the NOVX protein to bind to or interact with a NOVX target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (*i.e.* intracellular  $\text{Ca}^{2+}$ , diacylglycerol,  $\text{IP}_3$ , etc.), detecting catalytic/enzymatic activity of the target an appropriate substrate, detecting the  
10 induction of a reporter gene (comprising a NOVX-responsive regulatory element operatively linked to a nucleic acid encoding a detectable marker, *e.g.*, luciferase), or detecting a cellular response, for example, cell survival, cellular differentiation, or cell proliferation.

In yet another embodiment, an assay of the invention is a cell-free assay comprising contacting a NOVX protein or biologically-active portion thereof with a test compound and  
15 determining the ability of the test compound to bind to the NOVX protein or biologically-active portion thereof. Binding of the test compound to the NOVX protein can be determined either directly or indirectly as described above. In one such embodiment, the assay comprises contacting the NOVX protein or biologically-active portion thereof with a known compound which binds NOVX to form an assay mixture, contacting the assay mixture with a test  
20 compound, and determining the ability of the test compound to interact with a NOVX protein, wherein determining the ability of the test compound to interact with a NOVX protein comprises determining the ability of the test compound to preferentially bind to NOVX or biologically-active portion thereof as compared to the known compound.

In still another embodiment, an assay is a cell-free assay comprising contacting NOVX  
25 protein or biologically-active portion thereof with a test compound and determining the ability of the test compound to modulate (*e.g.* stimulate or inhibit) the activity of the NOVX protein or biologically-active portion thereof. Determining the ability of the test compound to modulate the activity of NOVX can be accomplished, for example, by determining the ability of the NOVX protein to bind to a NOVX target molecule by one of the methods described above for  
30 determining direct binding. In an alternative embodiment, determining the ability of the test compound to modulate the activity of NOVX protein can be accomplished by determining the ability of the NOVX protein further modulate a NOVX target molecule. For example, the

catalytic/enzymatic activity of the target molecule on an appropriate substrate can be determined as described above.

In yet another embodiment, the cell-free assay comprises contacting the NOVX protein or biologically-active portion thereof with a known compound which binds NOVX protein to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a NOVX protein, wherein determining the ability of the test compound to interact with a NOVX protein comprises determining the ability of the NOVX protein to preferentially bind to or modulate the activity of a NOVX target molecule.

The cell-free assays of the invention are amenable to use of both the soluble form or the membrane-bound form of NOVX protein. In the case of cell-free assays comprising the membrane-bound form of NOVX protein, it may be desirable to utilize a solubilizing agent such that the membrane-bound form of NOVX protein is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton<sup>®</sup> X-100, Triton<sup>®</sup> X-114, Thesit<sup>®</sup>, Isotridecypoly(ethylene glycol ether)<sub>n</sub>, N-dodecyl--N,N-dimethyl-3-ammonio-1-propane sulfonate, 3-(3-cholamidopropyl) dimethylamminiol-1-propane sulfonate (CHAPS), or 3-(3-cholamidopropyl)dimethylamminiol-2-hydroxy-1-propane sulfonate (CHAPSO).

In more than one embodiment of the above assay methods of the invention, it may be desirable to immobilize either NOVX protein or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to NOVX protein, or interaction of NOVX protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided that adds a domain that allows one or both of the proteins to be bound to a matrix. For example, GST-NOVX fusion proteins or GST-target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtiter plates, that are then combined with the test compound or the test compound and either the non-adsorbed target protein or NOVX protein, and the mixture is incubated under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any



unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described, *supra*. Alternatively, the complexes can be dissociated from the matrix, and the level of NOVX protein binding or activity determined using standard techniques.

5 Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either the NOVX protein or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated NOVX protein or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well-known within the art (*e.g.*, biotinylation kit, Pierce Chemicals, Rockford, Ill.), and  
10 immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with NOVX protein or target molecules, but which do not interfere with binding of the NOVX protein to its target molecule, can be derivatized to the wells of the plate, and unbound target or NOVX protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized  
15 complexes, include immunodetection of complexes using antibodies reactive with the NOVX protein or target molecule, as well as enzyme-linked assays that rely on detecting an enzymatic activity associated with the NOVX protein or target molecule.

In another embodiment, modulators of NOVX protein expression are identified in a method wherein a cell is contacted with a candidate compound and the expression of NOVX  
20 mRNA or protein in the cell is determined. The level of expression of NOVX mRNA or protein in the presence of the candidate compound is compared to the level of expression of NOVX mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of NOVX mRNA or protein expression based upon this comparison. For example, when expression of NOVX mRNA or protein is greater (*i.e.*, statistically  
25 significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of NOVX mRNA or protein expression. Alternatively, when expression of NOVX mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of NOVX mRNA or protein expression. The level of NOVX mRNA or protein  
30 expression in the cells can be determined by methods described herein for detecting NOVX mRNA or protein.

5 In yet another aspect of the invention, the NOVX proteins can be used as "bait proteins" in a two-hybrid assay or three hybrid assay (see, e.g., U.S. Patent No. 5,283,317; Zervos, *et al.*, 1993. *Cell* 72: 223-232; Madura, *et al.*, 1993. *J. Biol. Chem.* 268: 12046-12054; Bartel, *et al.*, 1993. *Biotechniques* 14: 920-924; Iwabuchi, *et al.*, 1993. *Oncogene* 8: 1693-1696; and Brent WO 94/10300), to identify other proteins that bind to or interact with NOVX ("NOVX-binding proteins" or "NOVX-bp") and modulate NOVX activity. Such NOVX-binding proteins are also likely to be involved in the propagation of signals by the NOVX proteins as, for example, upstream or downstream elements of the NOVX pathway.

10 The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for NOVX is fused to a gene encoding the DNA binding domain of a known transcription factor (e.g., GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming a NOVX-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (e.g., LacZ) that is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene that encodes the protein which interacts with NOVX.

The invention further pertains to novel agents identified by the aforementioned screening assays and uses thereof for treatments as described herein.

## 25 **Detection Assays**

Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. By way of example, and not of limitation, these sequences can be used to: (i) identify an individual from a minute biological sample (tissue typing); and (ii) aid in forensic identification of a biological sample. Some of these applications are described in the subsections, below.

## **Tissue Typing**

5 The NOVX sequences of the invention can be used to identify individuals from minute biological samples. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. The sequences of the invention are useful as additional DNA markers for RFLP ("restriction fragment length polymorphisms," described in U.S. Patent No. 5,272,057).

10 Furthermore, the sequences of the invention can be used to provide an alternative technique that determines the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the NOVX sequences described herein can be used to prepare two PCR primers from the 5'- and 3'-termini of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

15 Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the invention can be used to obtain such identification sequences from individuals and from tissue. The NOVX sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Much of the allelic variation is due to single nucleotide polymorphisms (SNPs), which include restriction fragment length polymorphisms (RFLPs).

20 Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers that each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences, such as those in SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23 or 25 are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

### **Predictive Medicine**

30 The invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, pharmacogenomics, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of

the invention relates to diagnostic assays for determining NOVX protein and/or nucleic acid expression as well as NOVX activity, in the context of a biological sample (*e.g.*, blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant NOVX expression or activity.

- 5 Disorders associated with aberrant NOVX expression of activity include, for example, disorders of olfactory loss, *e.g.* trauma, HIV illness, neoplastic growth, and neurological disorders, *e.g.* Parkinson's disease and Alzheimer's disease.

The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with NOVX protein, nucleic acid  
10 expression or activity. For example, mutations in a NOVX gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with NOVX protein, nucleic acid expression, or biological activity.

Another aspect of the invention provides methods for determining NOVX protein, nucleic  
15 acid expression or activity in an individual to thereby select appropriate therapeutic or prophylactic agents for that individual (referred to herein as "pharmacogenomics"). Pharmacogenomics allows for the selection of agents (*e.g.*, drugs) for therapeutic or prophylactic treatment of an individual based on the genotype of the individual (*e.g.*, the genotype of the individual examined to determine the ability of the individual to respond to a particular agent.)

20 Yet another aspect of the invention pertains to monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of NOVX in clinical trials.

These and other agents are described in further detail in the following sections.

### Diagnostic Assays

25 An exemplary method for detecting the presence or absence of NOVX in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting NOVX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) that encodes NOVX protein such that the presence of NOVX is detected in the biological sample. An agent for detecting NOVX mRNA or genomic DNA is a labeled  
30 nucleic acid probe capable of hybridizing to NOVX mRNA or genomic DNA. The nucleic acid probe can be, for example, a full-length NOVX nucleic acid, such as the nucleic acid of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23 or 25, or a portion thereof, such as an oligonucleotide

of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to NOVX mRNA or genomic DNA. Other suitable probes for use in the diagnostic assays of the invention are described herein.

One agent for detecting NOVX protein is an antibody capable of binding to NOVX protein, preferably an antibody with a detectable label. Antibodies directed against a protein of the invention may be used in methods known within the art relating to the localization and/or quantitation of the protein (e.g., for use in measuring levels of the protein within appropriate physiological samples, for use in diagnostic methods, for use in imaging the protein, and the like). In a given embodiment, antibodies against the proteins, or derivatives, fragments, analogs or homologs thereof, that contain the antigen binding domain, are utilized as pharmacologically-active compounds.

An antibody specific for a protein of the invention can be used to isolate the protein by standard techniques, such as immunoaffinity chromatography or immunoprecipitation. Such an antibody can facilitate the purification of the natural protein antigen from cells and of recombinantly produced antigen expressed in host cells. Moreover, such an antibody can be used to detect the antigenic protein (e.g., in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the antigenic protein. Antibodies directed against the protein can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, e.g., to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase,  $\beta$ -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include  $^{125}\text{I}$ ,  $^{131}\text{I}$ ,  $^{35}\text{S}$  or  $^3\text{H}$ .

Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (e.g., Fab or  $\text{F(ab')}_2$ ) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (i.e.,

physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently-labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently-labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect NOVX mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of NOVX mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of NOVX protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations, and immunofluorescence. *In vitro* techniques for detection of NOVX genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of NOVX protein include introducing into a subject a labeled anti-NOVX antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject.

In one embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting NOVX protein, mRNA, or genomic DNA, such that the presence of NOVX protein, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of NOVX protein, mRNA or genomic DNA in the control sample with the presence of NOVX protein, mRNA or genomic DNA in the test sample.

The invention also encompasses kits for detecting the presence of NOVX in a biological sample. For example, the kit can comprise: a labeled compound or agent capable of detecting NOVX protein or mRNA in a biological sample; means for determining the amount of NOVX in the sample; and means for comparing the amount of NOVX in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect NOVX protein or nucleic acid.





characterized by at least one of an alteration affecting the integrity of a gene encoding a NOVX-protein, or the misexpression of the NOVX gene. For example, such genetic lesions can be detected by ascertaining the existence of at least one of: (i) a deletion of one or more nucleotides from a NOVX gene; (ii) an addition of one or more nucleotides to a NOVX gene; (iii) a substitution of one or more nucleotides of a NOVX gene, (iv) a chromosomal rearrangement of a NOVX gene; (v) an alteration in the level of a messenger RNA transcript of a NOVX gene, (vi) aberrant modification of a NOVX gene, such as of the methylation pattern of the genomic DNA, (vii) the presence of a non-wild-type splicing pattern of a messenger RNA transcript of a NOVX gene, (viii) a non-wild-type level of a NOVX protein, (ix) allelic loss of a NOVX gene, and (x) inappropriate post-translational modification of a NOVX protein. As described herein, there are a large number of assay techniques known in the art which can be used for detecting lesions in a NOVX gene. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

In certain embodiments, detection of the lesion involves the use of a probe/primer in a polymerase chain reaction (PCR) (*see, e.g.*, U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (*see, e.g.*, Landegran, *et al.*, 1988. *Science* 241: 1077-1080; and Nakazawa, *et al.*, 1994. *Proc. Natl. Acad. Sci. USA* 91: 360-364), the latter of which can be particularly useful for detecting point mutations in the NOVX-gene (*see*, Abravaya, *et al.*, 1995. *Nucl. Acids Res.* 23: 675-682). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (*e.g.*, genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers that specifically hybridize to a NOVX gene under conditions such that hybridization and amplification of the NOVX gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (*see*, Guatelli, *et al.*, 1990. *Proc. Natl. Acad. Sci. USA* 87: 1874-1878), transcriptional amplification system (*see*, Kwoh, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA* 86: 1173-1177); Q $\beta$  Replicase (*see*, Lizardi, *et al.*, 1988. *BioTechnology* 6: 1197), or any other nucleic acid amplification method,

followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In an alternative embodiment, mutations in a NOVX gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (*see, e.g.,* U.S. Patent No. 5,493,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations in NOVX can be identified by hybridizing a sample and control nucleic acids, *e.g.,* DNA or RNA, to high-density arrays containing hundreds or thousands of oligonucleotides probes. *See, e.g.,* Cronin, *et al.*, 1996. *Human Mutation* 7: 244-255; Kozal, *et al.*, 1996. *Nat. Med.* 2: 753-759. For example, genetic mutations in NOVX can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin, *et al.*, *supra*. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the NOVX gene and detect mutations by comparing the sequence of the sample NOVX with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxim and Gilbert, 1977. *Proc. Natl. Acad. Sci. USA* 74: 560 or Sanger, 1977. *Proc. Natl. Acad. Sci. USA* 74: 5463. It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays (*see, e.g.,* Naeve, *et al.*, 1995. *Biotechniques* 19: 448), including sequencing by mass spectrometry (*see, e.g.,* PCT International Publication No. WO 94/16101;

Cohen, *et al.*, 1996. *Adv. Chromatography* 36: 127-162; and Griffin, *et al.*, 1993. *Appl. Biochem. Biotechnol.* 38: 147-159).

Other methods for detecting mutations in the NOVX gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes. *See, e.g.*, Myers, *et al.*, 1985. *Science* 230: 1242. In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes of formed by hybridizing (labeled) RNA or DNA containing the wild-type NOVX sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent that cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S<sub>1</sub> nuclease to enzymatically digesting the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. *See, e.g.*, Cotton, *et al.*, 1988. *Proc. Natl. Acad. Sci. USA* 85: 4397; Saleeba, *et al.*, 1992. *Methods Enzymol.* 217: 286-295. In an embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in NOVX cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches. *See, e.g.*, Hsu, *et al.*, 1994. *Carcinogenesis* 15: 1657-1662. According to an exemplary embodiment, a probe based on a NOVX sequence, *e.g.*, a wild-type NOVX sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. *See, e.g.*, U.S. Patent No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in NOVX genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids. *See, e.g.*, Orita, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA*: 86: 2766; Cotton, 1993. *Mutat.*

*Res.* 285: 125-144; Hayashi, 1992. *Genet. Anal. Tech. Appl.* 9: 73-79. Single-stranded DNA fragments of sample and control NOVX nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change.

5 The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In one embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility. *See, e.g., Keen, et al., 1991. Trends Genet.* 7: 5.

10 In yet another embodiment, the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE). *See, e.g., Myers, et al., 1985. Nature* 313: 495. When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by

15 PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA. *See, e.g., Rosenbaum and Reissner, 1987. Biophys. Chem.* 265: 12753.

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension.

20 For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions that permit hybridization only if a perfect match is found. *See, e.g., Saiki, et al., 1986. Nature* 324: 163; Saiki, et al., 1989. *Proc. Natl. Acad. Sci. USA* 86: 6230. Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached

25 to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology that depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization; *see, e.g., Gibbs, et al., 1989. Nucl.*

30 *Acids Res.* 17: 2437-2448) or at the extreme 3'-terminus of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (*see, e.g., Prossner, 1993. Tibtech.* 11: 238). In addition it may be desirable to introduce a novel restriction site in the

region of the mutation to create cleavage-based detection. See, e.g., Gasparini, *et al.*, 1992. *Mol. Cell Probes* 6: 1. It is anticipated that in certain embodiments amplification may also be performed using *Taq* ligase for amplification. See, e.g., Barany, 1991. *Proc. Natl. Acad. Sci. USA* 88: 189. In such cases, ligation will occur only if there is a perfect match at the 3'-terminus of the 5' sequence, making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, e.g., in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a NOVX gene.

Furthermore, any cell type or tissue, preferably peripheral blood leukocytes, in which NOVX is expressed may be utilized in the prognostic assays described herein. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

### **Pharmacogenomics**

Agents, or modulators that have a stimulatory or inhibitory effect on NOVX activity (e.g., NOVX gene expression), as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) disorders (e.g. disorders of olfactory loss, e.g. trauma, HIV illness, neoplastic growth, and neurological disorders, e.g. Parkinson's disease and Alzheimer's disease). In conjunction with such treatment, the pharmacogenomics (i.e., the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) of the individual may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, the pharmacogenomics of the individual permits the selection of effective agents (e.g., drugs) for prophylactic or therapeutic treatments based on a consideration of the individual's genotype. Such pharmacogenomics can further be used to determine appropriate dosages and therapeutic regimens. Accordingly, the activity of NOVX protein, expression of NOVX nucleic acid, or mutation content of NOVX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual.

Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See e.g.,



Eichelbaum, 1996. *Clin. Exp. Pharmacol. Physiol.*, 23: 983-985; Linder, 1997. *Clin. Chem.*, 43: 254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare defects or as polymorphisms. For example, glucose-6-phosphate dehydrogenase (G6PD) deficiency is a common inherited enzymopathy in which the main clinical complication is hemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (*e.g.*, N-acetyltransferase 2 (NAT 2) and cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, PM show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. At the other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Thus, the activity of NOVX protein, expression of NOVX nucleic acid, or mutation content of NOVX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual. In addition, pharmacogenetic studies can be used to apply genotyping of polymorphic alleles encoding drug-metabolizing enzymes to the identification of an individual's drug responsiveness phenotype. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus

enhance therapeutic or prophylactic efficiency when treating a subject with a NOVX modulator, such as a modulator identified by one of the exemplary screening assays described herein.

### Monitoring of Effects During Clinical Trials

5           Monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of NOVX (*e.g.*, the ability to modulate aberrant cell proliferation) can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase NOVX gene expression, protein levels, or upregulate NOVX activity, can be monitored in clinical trails of subjects exhibiting  
10   decreased NOVX gene expression, protein levels, or downregulated NOVX activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease NOVX gene expression, protein levels, or downregulate NOVX activity, can be monitored in clinical trails of subjects exhibiting increased NOVX gene expression, protein levels, or upregulated NOVX activity. In such clinical trials, the expression or activity of NOVX and, preferably, other  
15   genes that have been implicated in, for example, a cellular proliferation or immune disorder can be used as a "read out" or markers of the immune responsiveness of a particular cell.

          By way of example, and not of limitation, genes, including NOVX, that are modulated in cells by treatment with an agent (*e.g.*, compound, drug or small molecule) that modulates NOVX activity (*e.g.*, identified in a screening assay as described herein) can be identified. Thus, to  
20   study the effect of agents on cellular proliferation disorders, for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of NOVX and other genes implicated in the disorder. The levels of gene expression (*i.e.*, a gene expression pattern) can be quantified by Northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by  
25   measuring the levels of activity of NOVX or other genes. In this manner, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during, treatment of the individual with the agent.

          In one embodiment, the invention provides a method for monitoring the effectiveness of  
30   treatment of a subject with an agent (*e.g.*, an agonist, antagonist, protein, peptide, peptidomimetic, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) comprising the steps of (*i*) obtaining a pre-administration sample from a

subject prior to administration of the agent; (ii) detecting the level of expression of a NOVX protein, mRNA, or genomic DNA in the preadministration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression or activity of the NOVX protein, mRNA, or genomic DNA in the post-administration samples; (v) comparing the level of expression or activity of the NOVX protein, mRNA, or genomic DNA in the pre-administration sample with the NOVX protein, mRNA, or genomic DNA in the post administration sample or samples; and (vi) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of NOVX to higher levels than detected, *i.e.*, to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease expression or activity of NOVX to lower levels than detected, *i.e.*, to decrease the effectiveness of the agent.

#### Methods of Treatment

The invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant NOVX expression or activity. Disorders associated with aberrant NOVX expression include, for example, disorders of olfactory loss, *e.g.* trauma, HIV illness, neoplastic growth, and neurological disorders, *e.g.* Parkinson's disease and Alzheimer's disease.

These methods of treatment will be discussed more fully, below.

#### Disease and Disorders

Diseases and disorders that are characterized by increased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with Therapeutics that antagonize (*i.e.*, reduce or inhibit) activity. Therapeutics that antagonize activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized include, but are not limited to: (i) an aforementioned peptide, or analogs, derivatives, fragments or homologs thereof; (ii) antibodies to an aforementioned peptide; (iii) nucleic acids encoding an aforementioned peptide; (iv) administration of antisense nucleic acid and nucleic acids that are "dysfunctional" (*i.e.*, due to a heterologous insertion within the coding sequences of coding sequences to an aforementioned peptide) that are utilized to "knockout" endogenous function of an aforementioned peptide by homologous recombination (*see, e.g.*, Capecchi, 1989. *Science* 244: 1288-1292); or (v) modulators (*i.e.*, inhibitors, agonists and antagonists, including

additional peptide mimetic of the invention or antibodies specific to a peptide of the invention) that alter the interaction between an aforementioned peptide and its binding partner.

Diseases and disorders that are characterized by decreased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with

5 Therapeutics that increase (*i.e.*, are agonists to) activity. Therapeutics that upregulate activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized include, but are not limited to, an aforementioned peptide, or analogs, derivatives, fragments or homologs thereof; or an agonist that increases bioavailability.

10 Increased or decreased levels can be readily detected by quantifying peptide and/or RNA, by obtaining a patient tissue sample (*e.g.*, from biopsy tissue) and assaying it *in vitro* for RNA or peptide levels, structure and/or activity of the expressed peptides (or mRNAs of an aforementioned peptide). Methods that are well-known within the art include, but are not limited to, immunoassays (*e.g.*, by Western blot analysis, immunoprecipitation followed by sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis, immunocytochemistry, etc.) and/or  
15 hybridization assays to detect expression of mRNAs (*e.g.*, Northern assays, dot blots, *in situ* hybridization, and the like).

### Prophylactic Methods

20 In one aspect, the invention provides a method for preventing, in a subject, a disease or condition associated with an aberrant NOVX expression or activity, by administering to the subject an agent that modulates NOVX expression or at least one NOVX activity. Subjects at risk for a disease that is caused or contributed to by aberrant NOVX expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms  
25 characteristic of the NOVX aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending upon the type of NOVX aberrancy, for example, a NOVX agonist or NOVX antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein. The prophylactic methods of the invention are further discussed in the following subsections.

## Therapeutic Methods

Another aspect of the invention pertains to methods of modulating NOVX expression or activity for therapeutic purposes. The modulatory method of the invention involves contacting a cell with an agent that modulates one or more of the activities of NOVX protein activity associated with the cell. An agent that modulates NOVX protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring cognate ligand of a NOVX protein, a peptide, a NOVX peptidomimetic, or other small molecule. In one embodiment, the agent stimulates one or more NOVX protein activity. Examples of such stimulatory agents include active NOVX protein and a nucleic acid molecule encoding NOVX that has been introduced into the cell. In another embodiment, the agent inhibits one or more NOVX protein activity. Examples of such inhibitory agents include antisense NOVX nucleic acid molecules and anti-NOVX antibodies. These modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g., by administering the agent to a subject). As such, the invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant expression or activity of a NOVX protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination of agents that modulates (e.g., up-regulates or down-regulates) NOVX expression or activity. In another embodiment, the method involves administering a NOVX protein or nucleic acid molecule as therapy to compensate for reduced or aberrant NOVX expression or activity.

Stimulation of NOVX activity is desirable in situations in which NOVX is abnormally downregulated and/or in which increased NOVX activity is likely to have a beneficial effect. One example of such a situation is where a subject has a disorder characterized by aberrant cell proliferation and/or differentiation (e.g., cancer or immune associated). Another example of such a situation is where the subject has an immunodeficiency disease (e.g., AIDS).

Antibodies of the invention, including polyclonal, monoclonal, humanized and fully human antibodies, may be used as therapeutic agents. Such agents will generally be employed to treat or prevent a disease or pathology in a subject. An antibody preparation, preferably one having high specificity and high affinity for its target antigen, is administered to the subject and will generally have an effect due to its binding with the target. Such an effect may be one of two kinds, depending on the specific nature of the interaction between the given antibody molecule

and the target antigen in question. In the first instance, administration of the antibody may abrogate or inhibit the binding of the target with an endogenous ligand to which it naturally binds. In this case, the antibody binds to the target and masks a binding site of the naturally occurring ligand, wherein the ligand serves as an effector molecule. Thus the receptor mediates a signal transduction pathway for which ligand is responsible.

Alternatively, the effect may be one in which the antibody elicits a physiological result by virtue of binding to an effector binding site on the target molecule. In this case the target, a receptor having an endogenous ligand which may be absent or defective in the disease or pathology, binds the antibody as a surrogate effector ligand, initiating a receptor-based signal transduction event by the receptor.

A therapeutically effective amount of an antibody of the invention relates generally to the amount needed to achieve a therapeutic objective. As noted above, this may be a binding interaction between the antibody and its target antigen that, in certain cases, interferes with the functioning of the target, and in other cases, promotes a physiological response. The amount required to be administered will furthermore depend on the binding affinity of the antibody for its specific antigen, and will also depend on the rate at which an administered antibody is depleted from the free volume other subject to which it is administered. Common ranges for therapeutically effective dosing of an antibody or antibody fragment of the invention may be, by way of nonlimiting example, from about 0.1 mg/kg body weight to about 50 mg/kg body weight. Common dosing frequencies may range, for example, from twice daily to once a week.

#### **Determination of the Biological Effect of the Therapeutic**

In various embodiments of the invention, suitable *in vitro* or *in vivo* assays are performed to determine the effect of a specific Therapeutic and whether its administration is indicated for treatment of the affected tissue.

In various specific embodiments, *in vitro* assays may be performed with representative cells of the type(s) involved in the patient's disorder, to determine if a given Therapeutic exerts the desired effect upon the cell type(s). Compounds for use in therapy may be tested in suitable animal model systems including, but not limited to rats, mice, chicken, cows, monkeys, rabbits, and the like, prior to testing in human subjects. Similarly, for *in vivo* testing, any of the animal model system known in the art may be used prior to administration to human subjects.



The invention will be further described in the following examples, which do not limit the scope of the invention described in the claims.

## EXAMPLES

### 5        Example 1- Method of Identifying the Nucleic Acids Encoding the G-Protein Coupled Receptors.

Novel nucleic acid sequences were identified by TblastN using CuraGen Corporation's sequence file run against the Genomic Daily Files made available by GenBank. The nucleic acids were further predicted by the program GenScan™, including selection of exons. These  
10        were further modified by means of similarities using BLAST searches. The sequences were then manually corrected for apparent inconsistencies, thereby obtaining the sequences encoding the full-length protein.

### 15        Example 2 - Quantitative expression analysis of clones in various cells and tissues

The quantitative expression of various clones was assessed using microtiter plates containing RNA samples from a variety of normal and pathology-derived cells, cell lines and tissues using real time quantitative PCR (RTQ PCR; TAQMAN®). RTQ PCR was performed on a Perkin-Elmer Biosystems ABI PRISM® 7700 Sequence Detection System. Various collections of samples are assembled on the plates, and referred to as Panel 1 (containing cells  
20        and cell lines from normal and cancer sources), Panel 2 (containing samples derived from tissues, in particular from surgical samples, from normal and cancer sources), Panel 3 (containing samples derived from a wide variety of cancer sources), Panel 4 (containing cells and cell lines from normal cells and cells related to inflammatory conditions) and Panel CNSD.01 (containing samples from normal and diseased brains).

First, the RNA samples were normalized to constitutively expressed genes such as b-actin and GAPDH. RNA (~50 ng total or ~1 ng polyA+) was converted to cDNA using the TAQMAN® Reverse Transcription Reagents Kit (PE Biosystems, Foster City, CA; Catalog No. N808-0234) and random hexamers according to the manufacturer's protocol. Reactions were performed in 20 ul and incubated for 30 min. at 480C. cDNA (5 ul) was then transferred to a  
30        separate plate for the TAQMAN® reaction using b-actin and GAPDH TAQMAN® Assay Reagents (PE Biosystems; Catalog Nos. 4310881E and 4310884E, respectively) and TAQMAN® universal PCR Master Mix (PE Biosystems; Catalog No. 4304447) according to the manufacturer's protocol. Reactions were performed in 25 ul using the following parameters: 2

min. at 500C; 10 min. at 950C; 15 sec. at 950C/1 min. at 600C (40 cycles). Results were recorded as CT values (cycle at which a given sample crosses a threshold level of fluorescence) using a log scale, with the difference in RNA concentration between a given sample and the sample with the lowest CT value being represented as 2 to the power of delta CT. The percent  
5 relative expression is then obtained by taking the reciprocal of this RNA difference and multiplying by 100. The average CT values obtained for  $\beta$ -actin and GAPDH were used to normalize RNA samples. The RNA sample generating the highest CT value required no further diluting, while all other samples were diluted relative to this sample according to their b-actin /GAPDH average CT values.

10 Normalized RNA (5 ul) was converted to cDNA and analyzed via TAQMAN® using One Step RT-PCR Master Mix Reagents (PE Biosystems; Catalog No. 4309169) and gene-specific primers according to the manufacturer's instructions. Probes and primers were designed for each assay according to Perkin Elmer Biosystem's Primer Express Software package (version I for Apple Computer's Macintosh Power PC) or a similar algorithm using the target sequence as  
15 input. Default settings were used for reaction conditions and the following parameters were set before selecting primers: primer concentration = 250 nM, primer melting temperature ( $T_m$ ) range = 58°-60° C, primer optimal  $T_m$  = 59° C, maximum primer difference = 2° C, probe does not have 5' G, probe  $T_m$  must be 10° C greater than primer  $T_m$ , amplicon size 75 bp to 100 bp. The probes and primers selected (see below) were synthesized by Synthegen (Houston, TX, USA).  
20 Probes were double purified by HPLC to remove uncoupled dye and evaluated by mass spectroscopy to verify coupling of reporter and quencher dyes to the 5' and 3' ends of the probe, respectively. Their final concentrations were: forward and reverse primers, 900 nM each, and probe, 200nM.

PCR conditions: Normalized RNA from each tissue and each cell line was spotted in  
25 each well of a 96 well PCR plate (Perkin Elmer Biosystems). PCR cocktails including two probes (a probe specific for the target clone and another gene-specific probe multiplexed with the target probe) were set up using 1X TaqMan® PCR Master Mix for the PE Biosystems 7700, with 5 mM  $MgCl_2$ , dNTPs (dA, G, C, U at 1:1:1:2 ratios), 0.25 U/ml AmpliTaq Gold® (PE Biosystems), and 0.4 U/ml RNase inhibitor, and 0.25 U/ml reverse transcriptase. Reverse  
30 transcription was performed at 48° C for 30 minutes followed by amplification/PCR cycles as follows: 95° C 10 min, then 40 cycles of 95° C for 15 seconds, 60° C for 1 minute.

In the results for Panel 1, the following abbreviations are used:

ca. = carcinoma,

\* = established from metastasis,

met = metastasis,

s cell var= small cell variant,

non-s = non-sm =non-small,

squam = squamous,

pl. eff = pl effusion = pleural effusion,

glio = glioma,

astro = astrocytoma, and

neuro = neuroblastoma.

## Panel 2

The plates for Panel 2 generally include 2 control wells and 94 test samples composed of RNA or cDNA isolated from human tissue procured by surgeons working in close cooperation with the National Cancer Institute's Cooperative Human Tissue Network (CHTN) or the National Disease Research Initiative (NDRI). The tissues are derived from human malignancies and in cases where indicated many malignant tissues have "matched margins" obtained from noncancerous tissue just adjacent to the tumor. These are termed normal adjacent tissues and are denoted "NAT" in the results below. The tumor tissue and the "matched margins" are evaluated by two independent pathologists (the surgical pathologists and again by a pathologists at NDRI or CHTN). This analysis provides a gross histopathological assessment of tumor differentiation grade. Moreover, most samples include the original surgical pathology report that provides information regarding the clinical stage of the patient. These matched margins are taken from the tissue surrounding (i.e. immediately proximal) to the zone of surgery (designated "NAT", for normal adjacent tissue, in Table RR). In addition, RNA and cDNA samples were obtained from various human tissues derived from autopsies performed on elderly people or sudden death victims (accidents, etc.). These tissues were ascertained to be free of disease and were purchased from various commercial sources such as Clontech (Palo Alto, CA), Research Genetics, and Invitrogen.

RNA integrity from all samples is controlled for quality by visual assessment of agarose gel electropherograms using 28S and 18S ribosomal RNA staining intensity ratio as a guide (2:1 to 2.5:1 28s:18s) and the absence of low molecular weight RNAs that would be indicative of degradation products. Samples are controlled against genomic DNA contamination by RTQ PCR reactions run in the absence of reverse transcriptase using probe and primer sets designed to amplify across the span of a single exon.

#### Panel 3D

The plates of Panel 3D are comprised of 94 cDNA samples and two control samples. Specifically, 92 of these samples are derived from cultured human cancer cell lines, 2 samples of human primary cerebellar tissue and 2 controls. The human cell lines are generally obtained from ATCC (American Type Culture Collection), NCI or the German tumor cell bank and fall into the following tissue groups: Squamous cell carcinoma of the tongue, breast cancer, prostate cancer, melanoma, epidermoid carcinoma, sarcomas, bladder carcinomas, pancreatic cancers, kidney cancers, leukemias/lymphomas, ovarian/uterine/cervical, gastric, colon, lung and CNS cancer cell lines. In addition, there are two independent samples of cerebellum. These cells are all cultured under standard recommended conditions and RNA extracted using the standard procedures. The cell lines in panel 3D and 1.3D are of the most common cell lines used in the scientific literature.

RNA integrity from all samples is controlled for quality by visual assessment of agarose gel electropherograms using 28S and 18S ribosomal RNA staining intensity ratio as a guide (2:1 to 2.5:1 28s:18s) and the absence of low molecular weight RNAs that would be indicative of degradation products. Samples are controlled against genomic DNA contamination by RTQ PCR reactions run in the absence of reverse transcriptase using probe and primer sets designed to amplify across the span of a single exon.

#### Panel 4

Panel 4 includes samples on a 96 well plate (2 control wells, 94 test samples) composed of RNA (Panel 4r) or cDNA (Panel 4d) isolated from various human cell lines or tissues related to inflammatory conditions. Total RNA from control normal tissues such as colon and lung (Stratagene ,La Jolla, CA) and thymus and kidney (Clontech) were employed. Total RNA from liver tissue from cirrhosis patients and kidney from lupus patients was obtained from BioChain

(Biochain Institute, Inc., Hayward, CA). Intestinal tissue for RNA preparation from patients diagnosed as having Crohn's disease and ulcerative colitis was obtained from the National Disease Research Interchange (NDRI) (Philadelphia, PA).

5 Astrocytes, lung fibroblasts, dermal fibroblasts, coronary artery smooth muscle cells, small airway epithelium, bronchial epithelium, microvascular dermal endothelial cells, microvascular lung endothelial cells, human pulmonary aortic endothelial cells, human umbilical vein endothelial cells were all purchased from Clonetics (Walkersville, MD) and grown in the media supplied for these cell types by Clonetics. These primary cell types were activated with various cytokines or combinations of cytokines for 6 and/or 12-14 hours, as indicated. The  
10 following cytokines were used; IL-1 beta at approximately 1-5 ng/ml, TNF alpha at approximately 5-10 ng/ml, IFN gamma at approximately 20-50 ng/ml, IL-4 at approximately 5-10 ng/ml, IL-9 at approximately 5-10 ng/ml, IL-13 at approximately 5-10 ng/ml. Endothelial cells were sometimes starved for various times by culture in the basal media from Clonetics with 0.1% serum.

15 Mononuclear cells were prepared from blood of employees at CuraGen Corporation, using Ficoll. LAK cells were prepared from these cells by culture in DMEM 5% FCS (Hyclone), 100 mM non essential amino acids (Gibco/Life Technologies, Rockville, MD), 1 mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$  M (Gibco), and 10 mM Hepes (Gibco) and Interleukin 2 for 4-6 days. Cells were then either activated with 10-20 ng/ml PMA and 1-2  
20 mg/ml ionomycin, IL-12 at 5-10 ng/ml, IFN gamma at 20-50 ng/ml and IL-18 at 5-10 ng/ml for 6 hours. In some cases, mononuclear cells were cultured for 4-5 days in DMEM 5% FCS (Hyclone), 100 mM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$  M (Gibco), and 10 mM Hepes (Gibco) with PHA (phytohemagglutinin) or PWM (pokeweed mitogen) at approximately 5 mg/ml. Samples were  
25 taken at 24, 48 and 72 hours for RNA preparation. MLR (mixed lymphocyte reaction) samples were obtained by taking blood from two donors, isolating the mononuclear cells using Ficoll and mixing the isolated mononuclear cells 1:1 at a final concentration of approximately  $2 \times 10^6$  cells/ml in DMEM 5% FCS (Hyclone), 100 mM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol ( $5.5 \times 10^{-5}$  M) (Gibco), and 10 mM Hepes (Gibco).  
30 The MLR was cultured and samples taken at various time points ranging from 1- 7 days for RNA preparation.

Monocytes were isolated from mononuclear cells using CD14 Miltenyi Beads, +ve VS selection columns and a Vario Magnet according to the manufacturer's instructions. Monocytes were differentiated into dendritic cells by culture in DMEM 5% fetal calf serum (FCS) (Hyclone, Logan, UT), 100 mM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$  M (Gibco), and 10 mM Hepes (Gibco), 50 ng/ml GMCSF and 5 ng/ml IL-4 for 5-7 days. Macrophages were prepared by culture of monocytes for 5-7 days in DMEM 5% FCS (Hyclone), 100 mM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$  M (Gibco), 10 mM Hepes (Gibco) and 10% AB Human Serum or MCSF at approximately 50 ng/ml. Monocytes, macrophages and dendritic cells were stimulated for 6 and 12-14 hours with lipopolysaccharide (LPS) at 100 ng/ml. Dendritic cells were also stimulated with anti-CD40 monoclonal antibody (Pharmingen) at 10 mg/ml for 6 and 12-14 hours.

CD4 lymphocytes, CD8 lymphocytes and NK cells were also isolated from mononuclear cells using CD4, CD8 and CD56 Miltenyi beads, positive VS selection columns and a Vario Magnet according to the manufacturer's instructions. CD45RA and CD45RO CD4 lymphocytes were isolated by depleting mononuclear cells of CD8, CD56, CD14 and CD19 cells using CD8, CD56, CD14 and CD19 Miltenyi beads and positive selection. Then CD45RO beads were used to isolate the CD45RO CD4 lymphocytes with the remaining cells being CD45RA CD4 lymphocytes. CD45RA CD4, CD45RO CD4 and CD8 lymphocytes were placed in DMEM 5% FCS (Hyclone), 100 mM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$  M (Gibco), and 10 mM Hepes (Gibco) and plated at  $10^6$  cells/ml onto Falcon 6 well tissue culture plates that had been coated overnight with 0.5 mg/ml anti-CD28 (Pharmingen) and 3 ug/ml anti-CD3 (OKT3, ATCC) in PBS. After 6 and 24 hours, the cells were harvested for RNA preparation. To prepare chronically activated CD8 lymphocytes, we activated the isolated CD8 lymphocytes for 4 days on anti-CD28 and anti-CD3 coated plates and then harvested the cells and expanded them in DMEM 5% FCS (Hyclone), 100 mM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$  M (Gibco), and 10 mM Hepes (Gibco) and IL-2. The expanded CD8 cells were then activated again with plate bound anti-CD3 and anti-CD28 for 4 days and expanded as before. RNA was isolated 6 and 24 hours after the second activation and after 4 days of the second expansion culture. The isolated NK cells were cultured in DMEM 5% FCS (Hyclone), 100 mM non essential amino



acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$  M (Gibco), and 10 mM Hepes (Gibco) and IL-2 for 4-6 days before RNA was prepared.

To obtain B cells, tonsils were procured from NDRI. The tonsil was cut up with sterile dissecting scissors and then passed through a sieve. Tonsil cells were then spun down and resuspended at  $10^6$  cells/ml in DMEM 5% FCS (Hyclone), 100 mM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$  M (Gibco), and 10 mM Hepes (Gibco). To activate the cells, we used PWM at 5 mg/ml or anti-CD40 (Pharmingen) at approximately 10 mg/ml and IL-4 at 5-10 ng/ml. Cells were harvested for RNA preparation at 24, 48 and 72 hours.

To prepare the primary and secondary Th1/Th2 and Tr1 cells, six-well Falcon plates were coated overnight with 10  $\mu$ g/ml anti-CD28 (Pharmingen) and 2  $\mu$ g/ml OKT3 (ATCC), and then washed twice with PBS. Umbilical cord blood CD4 lymphocytes (Poietic Systems, German Town, MD) were cultured at  $10^5$ - $10^6$  cells/ml in DMEM 5% FCS (Hyclone), 100 mM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$  M (Gibco), 10 mM Hepes (Gibco) and IL-2 (4 ng/ml). IL-12 (5 ng/ml) and anti-IL4 (1  $\mu$ g/ml) were used to direct to Th1, while IL-4 (5 ng/ml) and anti-IFN gamma (1  $\mu$ g/ml) were used to direct to Th2 and IL-10 at 5 ng/ml was used to direct to Tr1. After 4-5 days, the activated Th1, Th2 and Tr1 lymphocytes were washed once in DMEM and expanded for 4-7 days in DMEM 5% FCS (Hyclone), 100 mM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$  M (Gibco), 10 mM Hepes (Gibco) and IL-2 (1 ng/ml). Following this, the activated Th1, Th2 and Tr1 lymphocytes were re-stimulated for 5 days with anti-CD28/OKT3 and cytokines as described above, but with the addition of anti-CD95L (1  $\mu$ g/ml) to prevent apoptosis. After 4-5 days, the Th1, Th2 and Tr1 lymphocytes were washed and then expanded again with IL-2 for 4-7 days. Activated Th1 and Th2 lymphocytes were maintained in this way for a maximum of three cycles. RNA was prepared from primary and secondary Th1, Th2 and Tr1 after 6 and 24 hours following the second and third activations with plate bound anti-CD3 and anti-CD28 mAbs and 4 days into the second and third expansion cultures in Interleukin 2.

The following leukocyte cells lines were obtained from the ATCC: Ramos, EOL-1, KU-812. EOL cells were further differentiated by culture in 0.1 mM dbcAMP at  $5 \times 10^5$  cells/ml for 8 days, changing the media every 3 days and adjusting the cell concentration to  $5 \times 10^5$  cells/ml. For the culture of these cells, we used DMEM or RPMI (as recommended by the ATCC), with

the addition of 5% FCS (Hyclone), 100 mM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$  M (Gibco), 10 mM Hepes (Gibco). RNA was either prepared from resting cells or cells activated with PMA at 10 ng/ml and ionomycin at 1 mg/ml for 6 and 14 hours. Keratinocyte line CCD106 and an airway epithelial tumor line NCI-H292 were also obtained from the ATCC. Both were cultured in DMEM 5% FCS (Hyclone), 100 mM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$  M (Gibco), and 10 mM Hepes (Gibco). CCD106 cells were activated for 6 and 14 hours with approximately 5 ng/ml TNF alpha and 1 ng/ml IL-1 beta, while NCI-H292 cells were activated for 6 and 14 hours with the following cytokines: 5 ng/ml IL-4, 5 ng/ml IL-9, 5 ng/ml IL-13 and 25 ng/ml IFN gamma.

For these cell lines and blood cells, RNA was prepared by lysing approximately  $10^7$  cells/ml using Trizol (Gibco BRL). Briefly, 1/10 volume of bromochloropropane (Molecular Research Corporation) was added to the RNA sample, vortexed and after 10 minutes at room temperature, the tubes were spun at 14,000 rpm in a Sorvall SS34 rotor. The aqueous phase was removed and placed in a 15 ml Falcon Tube. An equal volume of isopropanol was added and left at -20 degrees C overnight. The precipitated RNA was spun down at 9,000 rpm for 15 min in a Sorvall SS34 rotor and washed in 70% ethanol. The pellet was redissolved in 300 µl of RNase-free water and 35 µl buffer (Promega) 5 µl DTT, 7 µl RNAsin and 8 µl DNase were added. The tube was incubated at 37 degrees C for 30 minutes to remove contaminating genomic DNA, extracted once with phenol chloroform and re-precipitated with 1/10 volume of 3 M sodium acetate and 2 volumes of 100% ethanol. The RNA was spun down and placed in RNase free water. RNA was stored at -80 degrees C.

#### Panel CNSD.01

The plates for Panel CNSD.01 include two control wells and 94 test samples comprised of cDNA isolated from postmortem human brain tissue obtained from the Harvard Brain Tissue Resource Center. Brains are removed from calvaria of donors between 4 and 24 hours after death, sectioned by neuroanatomists, and frozen at -80°C in liquid nitrogen vapor. All brains are sectioned and examined by neuropathologists to confirm diagnoses with clear associated neuropathology.

Disease diagnoses are taken from patient records. The panel contains two brains from each of the following diagnoses: Alzheimer's disease, Parkinson's disease, Huntington's disease,

Progressive Supranuclear Palsy, Depression, and “Normal controls”. Within each of these brains, the following regions are represented: cingulate gyrus, temporal pole, globus palladus, substantia nigra, Brodman Area 4 (primary motor strip), Brodman Area 7 (parietal cortex), Brodman Area 9 (prefrontal cortex), and Brodman area 17 (occipital cortex). Not all brain regions are represented in all cases; e.g., Huntington’s disease is characterized in part by neurodegeneration in the globus palladus, thus this region is impossible to obtain from confirmed Huntington’s cases. Likewise Parkinson’s disease is characterized by degeneration of the substantia nigra making this region more difficult to obtain. Normal control brains were examined for neuropathology and found to be free of any pathology consistent with neurodegeneration.

RNA integrity from all samples is controlled for quality by visual assessment of agarose gel electropherograms using 28S and 18S ribosomal RNA staining intensity ratio as a guide (2:1 to 2.5:1 28s:18s) and the absence of low molecular weight RNAs that would be indicative of degradation products. Samples are controlled against genomic DNA contamination by RTQ PCR reactions run in the absence of reverse transcriptase using probe and primer sets designed to amplify across the span of a single exon.

In the labels employed to identify tissues in the CNS panel, the following abbreviations are used:

PSP = Progressive supranuclear palsy  
Sub Nigra = Substantia nigra  
Glob Palladus= Globus palladus  
Temp Pole = Temporal pole  
Cing Gyr = Cingulate gyrus  
BA 4 = Brodman Area 4

#### Example 2A: NOV3 Gene AL121986A

Expression of NOV3 (gene AL121986A) was assessed using the primer-probe set Ag295, Ag1628, and Ag2436, described in Tables 57-59. Results of the RTQ-PCR runs are shown in Tables 60, 61 and 62.

Table 57. Probe Name Ag295

Primers	Sequences	M	Length	Start Position
Forward	5'-GTGCCACCCAGCTGTTCTTT-3'		20	307
Probe	TET-5'-TTGGCTTTGCTTGACCAACTGCC- 3'-TAMRA		24	331
Reverse	5'-CGATCATATCCCATCACAGCAA-3'		22	361

Table 58. Probe Name Ag1628

Primers	Sequences	M	Length	Start Position
Forward	5'-GACATGGCACCTGTTATCAAGT-3'	9	22	543
Probe	TET-5'-CCTGCACTGACACCCATGTGAAAGAG- 3'-TAMRA	0	26	568
Reverse	5'-GGATGCTGAGGCTAAATAAAGC-3'	9.4	22	597

Table 59. Probe Name Ag2436

Primers	Sequences	M	Length	Start Position
Forward	5'-GGTGGCAGTGACCTACACA-3'	8	19	819
Probe	FAM-5'-TCCTCTTGTCTACAGTCTGAGGAACAA- 3'-TAMRA	3.8	27	858
Reverse	5'-CCAAGAACTCTTTTCAATGCA-3'	8	21	897

Table 60. Panel 1.2

Tissue Name	Relative Expression(%) )	Tissue Name	Relative Expression(%) )
	1.2tm1025t_ ag295		1.2tm1025t_ ag295
Endothelial cells	0.5	Renal ca. 786-0	3.2
Endothelial cells (treated)	0.5	Renal ca. A498	0.6
Pancreas	1.5	Renal ca. RXF 393	0.8
Pancreatic ca. CAPAN 2	0.5	Renal ca. ACHN	0.3
Adrenal Gland (new lot*)	1.3	Renal ca. UO-31	0.5
Thyroid	0.5	Renal ca. TK-10	0.8
Salivary gland	0.2	Liver	0.4
Pituitary gland	0.2	Liver (fetal)	1.6
Brain (fetal)	1.8	Liver ca. (hepatoblast) HepG2	0.3
Brain (whole)	1.3	Lung	0.6
Brain (amygdala)	1.0	Lung (fetal)	19.6
Brain (cerebellum)	1.4	Lung ca. (small cell) LX-1	1.0
Brain (hippocampus)	1.6	Lung ca. (small cell) NCI-H69	2.8
Brain (thalamus)	1.1	Lung ca. (s.cell var.) SHP-77	2.6
Cerebral Cortex	2.0	Lung ca. (large cell)NCI-H460	2.4
Spinal cord	0.4	Lung ca. (non-sm. cell) A549	0.6
CNS ca. (glio/astro) U87-MG	0.2	Lung ca. (non-s.cell) NCI-H23	2.6
CNS ca. (glio/astro) U-118-MG	0.4	Lung ca (non-s.cell) HOP-62	0.6
CNS ca. (astro) SW1783	0.6	Lung ca. (non-s.cl) NCI-H522	0.6

CNS ca.* (neuro; met ) SK-N-AS	0.2	Lung ca. (squam.) SW 900	0.4
CNS ca. (astro) SF-539	0.8	Lung ca. (squam.) NCI-H596	2.5
CNS ca. (astro) SNB-75	0.3	Mammary gland	2.0
		Breast ca.* (pl. effusion) MCF-7	1.6
CNS ca. (glio) SNB-19	0.7	Breast ca.* (pl.ef) MDA-MB-231	0.4
CNS ca. (glio) U251	0.9	Breast ca.* (pl. effusion) T47D	1.2
CNS ca. (glio) SF-295	0.2	Breast ca. BT-549	1.0
Heart	1.2	Breast ca. MDA-N	0.4
Skeletal Muscle (new lot*)	0.2	Ovary	0.3
Bone marrow	2.4	Ovarian ca. OVCAR-3	0.1
Thymus	100.0	Ovarian ca. OVCAR-4	0.1
Spleen	0.1	Ovarian ca. OVCAR-5	2.8
Lymph node	0.6	Ovarian ca. OVCAR-8	0.1
Colorectal	0.9	Ovarian ca. IGROV-1	0.1
Stomach	0.3	Ovarian ca.* (ascites) SK-OV-3	0.2
Small intestine	0.4	Uterus	0.2
Colon ca. SW480	0.1	Placenta	0.3
Colon ca.* (SW480 met)SW620	0.2	Prostate	0.4
Colon ca. HT29	0.4	Prostate ca.* (bone met)PC-3	0.8
Colon ca. HCT-116	0.0	Testis	29.5
Colon ca. CaCo-2	0.3		
<u>83219 CC Well to Mod</u>			
<u>Diff (ODO3866)</u>	4.2	Melanoma Hs688(A).T	1.4
Colon ca. HCC-2998	1.4	Melanoma* (met) Hs688(B).T	1.0



Gastric ca.* (liver met)			
NCI-N87	0.2	Melanoma UACC-62	1.7
Bladder	1.4	Melanoma M14	2.1
Trachea	0.9	Melanoma LOX IMVI	0.8
Kidney	0.2	Melanoma* (met) SK-MEL-5	2.2
Kidney (fetal)	1.8	Adipose	8.2

Table 61. Panel 1.3D

	Relative Expression(% )	Relative Expression(% )	Relative Expression(% )
	1.3Dt m3258t_ag295	1.3dx4tm5589 _ag1628_b1	1.3dx4tm5645 f_ag2436_a2
Tissue Name			
Liver adenocarcinoma	0.0	0.0	0.0
Pancreas	0.0	0.0	0.0
Pancreatic ca. CAPAN 2	2.9	0.0	0.0
Adrenal gland	3.4	0.0	0.0
Thyroid	0.0	0.0	0.0
Salivary gland	0.0	0.0	0.0
Pituitary gland	0.0	0.0	0.0
Brain (fetal)	0.0	0.0	0.0
Brain (whole)	0.0	0.0	0.0
Brain (amygdala)	0.0	0.0	0.0
Brain (cerebellum)	0.0	0.0	0.0
Brain (hippocampus)	0.0	0.0	0.0
Brain (substantia nigra)	0.0	0.0	0.0

Brain (thalamus)	0.0	0.0	0.0
Cerebral Cortex	0.0	0.0	0.0
Spinal cord	0.0	0.0	0.0
CNS ca. (glio/astro) U87-MG	0.0	0.0	0.0
CNS ca. (glio/astro) U-118-MG	0.0	0.0	0.0
CNS ca. (astro) SW1783	0.0	0.0	0.0
CNS ca.* (neuro; met ) SK-N-AS	0.0	0.0	0.0
CNS ca. (astro) SF-539	0.0	0.0	0.0
CNS ca. (astro) SNB-75	2.5	0.0	0.0
CNS ca. (glio) SNB-19	2.5	0.0	0.0
CNS ca. (glio) U251	0.0	0.0	0.0
CNS ca. (glio) SF-295	0.0	0.0	0.0
Heart (fetal)	0.0	0.0	0.0
Heart	0.0	0.0	0.0
Fetal Skeletal	1.9	0.0	0.0
Skeletal muscle	0.0	0.0	0.0
Bone marrow	0.0	0.0	0.0
Thymus	<b>100.0</b>	61.0	<b>100.0</b>
Spleen	0.0	<b>100.0</b>	0.0
Lymph node	0.0	0.0	0.0
Colorectal	13.8	0.0	0.0
Stomach	5.2	0.0	0.0
Small intestine	0.0	0.0	0.0
Colon ca. SW480	0.0	0.0	0.0
Colon ca.* (SW480 met)SW620	0.0	0.0	0.0
Colon ca. HT29	2.1	0.0	0.0



Colon ca. HCT-116	0.0	0.0	0.0
Colon ca. CaCo-2	0.0	0.0	0.0
<u>83219 CC Well to Mod Diff (ODO3866)</u>	0.0	0.0	0.0
Colon ca. HCC-2998	0.0	0.0	0.0
Gastric ca.* (liver met) NCI-N87	0.0	0.0	0.0
Bladder	0.0	0.0	0.0
Trachea	0.0	0.0	0.0
Kidney	0.0	0.0	0.0
Kidney (fetal)	0.0	0.0	0.0
Renal ca. 786-0	0.0	15.6	0.0
Renal ca. A498	0.0	0.0	0.0
Renal ca. RXF 393	0.0	6.6	0.0
Renal ca. ACHN	0.0	0.0	0.0
Renal ca. UO-31	0.0	0.0	0.0
Renal ca. TK-10	0.0	0.0	0.0
Liver	10.2	0.0	0.0
Liver (fetal)	0.0	0.0	0.0
Liver ca. (hepatoblast) HepG2	0.0	0.0	0.0
Lung	0.0	0.0	0.0
Lung (fetal)	0.0	0.0	0.0
Lung ca. (small cell) LX-1	0.0	0.0	0.0
Lung ca. (small cell) NCI-H69	0.0	0.0	0.0
Lung ca. (s.cell var.) SHP-77	0.0	0.0	0.0
Lung ca. (large cell)NCI-H460	3.6	0.0	0.0
Lung ca. (non-sm. cell) A549	0.0	0.0	0.0
Lung ca. (non-s.cell) NCI-H23	0.0	0.0	0.0

Lung ca (non-s.cell) HOP-62	0.0	0.0	0.0
Lung ca. (non-s.cl) NCI-H522	0.0	0.0	0.0
Lung ca. (squam.) SW 900	0.0	0.0	4.6
Lung ca. (squam.) NCI-H596	0.0	5.7	0.0
Mammary gland	0.0	0.0	0.0
Breast ca.* (pl. effusion) MCF-7	10.6	15.5	2.5
Breast ca.* (pl.ef) MDA-MB-231	0.0	0.0	0.0
Breast ca.* (pl. effusion) T47D	0.0	0.0	0.0
Breast ca. BT-549	4.0	0.0	0.0
Breast ca. MDA-N	0.0	0.0	0.0
Ovary	0.0	0.0	0.0
Ovarian ca. OVCAR-3	0.0	0.0	0.0
Ovarian ca. OVCAR-4	0.0	0.0	0.0
Ovarian ca. OVCAR-5	0.0	0.0	0.0
Ovarian ca. OVCAR-8	0.0	0.0	0.0
Ovarian ca. IGROV-1	2.7	0.0	0.0
Ovarian ca.* (ascites) SK-OV-3	3.2	15.4	0.0
Uterus	0.0	0.0	0.0
Placenta	0.0	0.0	0.0
Prostate	0.0	0.0	0.0
Prostate ca.* (bone met)PC-3	0.0	0.0	0.0
Testis	3.4	0.0	0.0
Melanoma Hs688(A).T	0.0	0.0	0.0
Melanoma* (met) Hs688(B).T	0.0	0.0	0.0
Melanoma UACC-62	0.0	0.0	0.0
Melanoma M14	0.0	0.0	0.0

Melanoma LOX IMVI	0.0	0.0	0.0
Melanoma* (met) SK-MEL-5	0.0	0.0	0.0
Adipose	0.0	0.0	0.0

Table 62. Panel 4D

Tissue Name	Relative Expression(% )	Relative Expression(% )
	4dx4tm5519t_ ag1628_a1	4dtm5005f_ ag2436
93768_Secondary Th1_anti-CD28/anti-CD3	0.0	0.0
93769_Secondary Th2_anti-CD28/anti-CD3	0.0	0.0
93770_Secondary Tr1_anti-CD28/anti-CD3	0.0	0.0
93573_Secondary Th1_resting day 4-6 in IL-2	0.0	0.0
93572_Secondary Th2_resting day 4-6 in IL-2	0.0	0.0
93571_Secondary Tr1_resting day 4-6 in IL-2	0.0	0.0
93568_primary Th1_anti-CD28/anti-CD3	0.0	0.0
93569_primary Th2_anti-CD28/anti-CD3	0.0	0.0
93570_primary Tr1_anti-CD28/anti-CD3	0.0	0.0
93565_primary Th1_resting dy 4-6 in IL-2	0.0	0.0
93566_primary Th2_resting dy 4-6 in IL-2	0.5	0.0
93567_primary Tr1_resting dy 4-6 in IL-2	0.0	0.0
93351_CD45RA CD4 lymphocyte_anti-CD28/anti- CD3	0.0	0.0
93352_CD45RO CD4 lymphocyte_anti-CD28/anti- CD3	0.0	0.0

93251_CD8 Lymphocytes_anti-CD28/anti-CD3	0.0	0.0
93353_chronic CD8 Lymphocytes 2ry_resting dy 4-6 in IL-2	0.0	0.0
93574_chronic CD8 Lymphocytes 2ry_activated CD3/CD28	0.0	0.0
93354_CD4_none	0.0	0.0
93252_Secondary Th1/Th2/Tr1_anti-CD95 CH11	0.0	0.0
93103_LAK cells_resting	0.0	0.0
93788_LAK cells_IL-2	0.0	0.0
93787_LAK cells_IL-2+IL-12	0.0	0.0
93789_LAK cells_IL-2+IFN gamma	0.0	0.0
93790_LAK cells_IL-2+ IL-18	0.0	0.0
93104_LAK cells_PMA/ionomycin and IL-18	0.0	0.0
93578_NK Cells IL-2_resting	0.0	0.0
93109_Mixed Lymphocyte Reaction_Two Way MLR	0.0	0.0
93110_Mixed Lymphocyte Reaction_Two Way MLR	0.0	0.0
93111_Mixed Lymphocyte Reaction_Two Way MLR	0.0	0.0
93112_Mononuclear Cells (PBMCs)_resting	0.0	0.0
93113_Mononuclear Cells (PBMCs)_PWM	0.0	0.0
93114_Mononuclear Cells (PBMCs)_PHA-L	0.0	0.0
93249_Ramos (B cell)_none	0.0	0.0
93250_Ramos (B cell)_ionomycin	0.0	0.0
93349_B lymphocytes_PWM	0.0	0.0
93350_B lymphocytes_CD40L and IL-4	0.0	0.0
92665_EOL-1 (Eosinophil)_dbcAMP differentiated	0.5	0.0
93248_EOL-1 (Eosinophil)_dbcAMP/PMAionomycin	0.3	0.0
93356_Dendritic Cells_none	0.0	0.0
93355_Dendritic Cells_LPS 100 ng/ml	0.0	0.0



93775_Dendritic Cells_anti-CD40	0.0	0.0
93774_Monocytes_resting	0.0	0.0
93776_Monocytes_LPS 50 ng/ml	0.0	0.0
93581_Macrophages_resting	0.0	0.0
93582_Macrophages_LPS 100 ng/ml	0.0	0.0
93098_HUVEC (Endothelial)_none	0.0	0.0
93099_HUVEC (Endothelial)_starved	0.0	0.0
93100_HUVEC (Endothelial)_IL-1b	0.0	0.0
93779_HUVEC (Endothelial)_IFN gamma	0.0	0.0
93102_HUVEC (Endothelial)_TNF alpha + IFN gamma	0.0	0.0
93101_HUVEC (Endothelial)_TNF alpha + IL4	0.0	0.0
93781_HUVEC (Endothelial)_IL-11	0.0	0.0
93583_Lung Microvascular Endothelial Cells_none	0.0	0.0
93584_Lung Microvascular Endothelial Cells_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0	0.0
92662_Microvascular Dermal endothelium_none	0.0	0.0
92663_Microsvascular Dermal endothelium_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0	0.0
93773_Bronchial epithelium_TNFa (4 ng/ml) and IL1b (1 ng/ml) **	0.0	0.0
93347_Small Airway Epithelium_none	0.0	0.0
93348_Small Airway Epithelium_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0	0.0
92668_Coronary Artery SMC_resting	0.0	0.0
92669_Coronary Artery SMC_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0	0.0
93107_astrocytes_resting	0.0	0.0
93108_astrocytes_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0	0.0

92666_KU-812 (Basophil)_resting	0.0	0.0
92667_KU-812 (Basophil)_PMA/ionoycin	0.0	0.0
93579_CCD1106 (Keratinocytes)_none	0.0	0.0
93580_CCD1106 (Keratinocytes)_TNFa and IFNg **	0.0	0.0
93791_Liver Cirrhosis	13.4	1.2
93792_Lupus Kidney	0.0	0.0
93577_NCI-H292	0.0	0.0
93358_NCI-H292_IL-4	0.0	0.0
93360_NCI-H292_IL-9	0.0	0.0
93359_NCI-H292_IL-13	0.0	0.7
93357_NCI-H292_IFN gamma	0.0	0.0
93777_HPAEC_-	0.0	0.0
93778_HPAEC_IL-1 beta/TNA alpha	0.0	0.0
93254_Normal Human Lung Fibroblast_none	0.0	0.0
93253_Normal Human Lung Fibroblast_TNFa (4 ng/ml) and IL-1b (1 ng/ml)	0.0	0.0
93257_Normal Human Lung Fibroblast_IL-4	0.0	0.0
93256_Normal Human Lung Fibroblast_IL-9	0.0	0.0
93255_Normal Human Lung Fibroblast_IL-13	0.0	0.0
93258_Normal Human Lung Fibroblast_IFN gamma	0.0	0.0
93106_Dermal Fibroblasts CCD1070_resting	0.0	0.0
93361_Dermal Fibroblasts CCD1070_TNF alpha 4 ng/ml	0.0	0.0
93105_Dermal Fibroblasts CCD1070_IL-1 beta 1 ng/ml	0.0	0.0
93772_dermal fibroblast_IFN gamma	0.0	0.0
93771_dermal fibroblast_IL-4	0.0	0.0
93259_IBD Colitis 1**	100.0	0.7
93260_IBD Colitis 2	0.9	0.0

93261_IBD Crohns	0.0	0.0
735010_Colon_normal	0.2	0.0
735019_Lung_none	0.0	0.0
64028-1_Thymus_none	0.0	0.0
64030-1_Kidney_none	39.4	100.0

### Summary of Panel Results:

5 Probe Ag295 in Panel 1.2 indicates expression of NOV3 (the AL121986A gene) as generally associated with normal tissues and with highest expression in thymus tissue. The gene is also moderately expressed in testis and fetal liver. Results with adipose tissue are unclear due to the possibility of contamination. Thus, the expression of this gene may be utilized to distinguish thymic tissue from other tissues.

10 In Panel 1.3D, probe Ag295/Ag1628/Ag2436 shows that AL121986A gene expression is consistent for thymus tissue in all three experiments using Panel 1.3D as it was in Panel 1.2, although no expression is seen in the thymus in panel 4 (Ag1628). The AL121986A gene or the protein encoded by this gene thus may be used as a marker for thymic tissue. Antibodies raised against the protein encoded by the AL121986A gene may be used as a tool to identify thymic  
15 tissue.

Panels 2D/2.2 with probes Ag295/Ag1628/Ag2436 indicate low to undetectable levels of expression of NOV3 (the AL121986A gene).

In Panel 4D, probes Ag1628/Ag2436 indicates expression of the AL121986A transcript in colitis 1, and at much lower levels in colitis 2. The protein encoded by the AL121986A gene  
20 may therefore be important in the inflammatory process during colitis. Antagonistic antibodies or small molecule therapeutic agents may reduce or inhibit inflammation in the bowel due to IBD. Probe Ag2436 shows expression of the gene in kidney tissues, but not in colitis or other samples.

## OTHER EMBODIMENTS

While the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages, and  
5 modifications are within the scope of the following claims.